

PCT

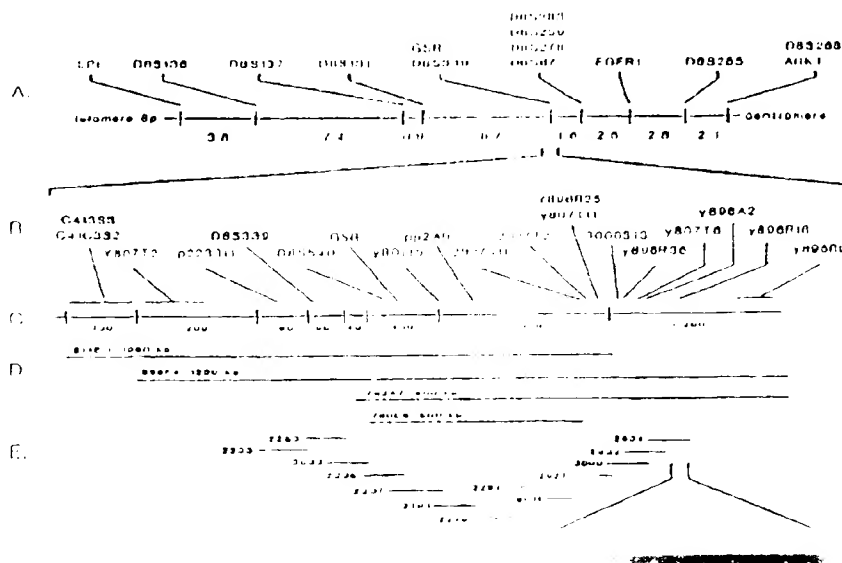
WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/00, 15/61, 9/90, 15/85, 15/86, 5/10, 5/20, C07K 16/40, C12Q 1/68, A01K 67/027, G01N 33/50		A1	(11) International Publication Number: WO 97/24435
(21) International Application Number: PCT/US96/20785		(43) International Publication Date: 10 July 1997 (10.07.97)	
(22) International Filing Date: 30 December 1996 (30.12.96)		(74) Agents: NOTTENBURG, Carol, Ph., D. et al.; Seed and Berry L.L.P., 6300 Columbia Center, 701 Fifth Avenue, Seattle, WA 98104-7092 (US).	
(30) Priority Data: 60/009,409 29 December 1995 (29.12.95) US 08/580,539 29 December 1995 (29.12.95) US 60/010,835 30 January 1996 (30.01.96) US 08/594,242 30 January 1996 (30.01.96) US 08/632,175 12 April 1996 (12.04.96) US		(81) Designated States: AL, AM, AT, AU, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).	
(71) Applicant: DARWIN MOLECULAR CORPORATION [US/US]; 1631-220th Street Southeast, Bothell, WA 98071 (US).		Published With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.	
(71)(72) Applicant and Inventor: OSHIMA, Junko [US/US]; 5200 Ravenna Avenue N.E., Seattle, WA 98105 (US).			
(72) Inventors: FU, Ying-Hui; 7417-28th Avenue N.W., Seattle, WA 98117 (US); YU, Chang-En; 5709-60th Avenue N.E., Seattle, WA 98105 (US); MULLIGAN, John; 5823-17th Avenue Northeast, Seattle, WA 98105 (US); SCHELLEN- BERG, Gerald, D.; 7031-19th Avenue Northwest, Seattle, WA 98117 (US).			

(54) Title: GENES AND GENE PRODUCTS RELATED TO WERNER'S SYNDROME



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic			SE	Sweden
CG	Congo	KR	Republic of Korea	SG	Singapore
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Ivory Coast				
CM	Cameroon				
CO	Colombia				
CR	Costa Rica				
CU	Cuba				
CY	Cyprus				
CZ	Czech Republic				
DE	Germany	LV	Latvia	TJ	Tajikistan
DK	Denmark	MC	Monaco	TT	Trinidad and Tobago
EE	Estonia	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	UG	Uganda
FI	Finland	ML	Mali	US	United States of America
FR	France	MS	Malta		
		MT	Maldives		

DESCRIPTION

GENES AND GENE PRODUCTS RELATED TO WERNER'S SYNDROME

5 TECHNICAL FIELD

The present invention relates generally to Werner's Syndrome and more specifically to methods and compositions suitable for use in diagnosis and treatment of Werner's Syndrome.

10 BACKGROUND OF THE INVENTION

Werner Syndrome (WS) is an autosomal recessive disorder with a complex phenotype. The disorder manifests itself in premature occurrence of age-related diseases and premature appearance of some of the physical features of normal aging. The onset of symptoms usually occurs after adolescence. The disorder progresses throughout life and typically patients have a shortened life expectancy with a modal age of death at 47. The prevalence of Werner Syndrome is estimated for heterozygotes to be 1-5 per 1,000 individuals, and for homozygotes to be 1-22 per 1,000,000 individuals.

Clinical symptoms of Werner Syndrome include both a prevalence of age-related diseases and physical features of aging. Such diseases include arteriosclerosis and heart disease, both benign and malignant neoplasms (usually sarcomas), diabetes mellitus, osteoporosis, and ocular cataracts. The physical appearance of WS patients is often manifest as a short stature, premature graying or loss of hair, hypogonadism, altered skin pigmentation, hyperkeratosis, tight skin, bird-like facies, cutaneous atrophy, cutaneous leg ulcers, and telangiectasia. Most of these diseases and features are present in from 40-90% of WS patients. Diagnosis of WS relies mainly upon the appearance of a certain number of these diseases and features

In addition to the noted signs and symptoms of aging, Werner Syndrome mimics normal aging as evidenced by the replicative potential of fibroblasts isolated from WS subjects. Replication potential of fibroblasts is reduced in these patients compared to fibroblasts isolated from age-matched controls, and is comparable to the replicative potential of fibroblasts taken from elderly subjects. Moreover, an increased mutation rate has been described in WS patients. Such abnormality is manifest as chromosomal instability, such as inversions, reciprocal translocations, deletions, and pseudodiploidy, and as increased mutation rate at the hypoxanthine phosphoribosyl transferase (HPRT) gene.

Werner Syndrome has been recognized as an autosomal recessive disorder. Goto et al. (Goto et al., *Nature* 355:735-738, 1992) mapped the WS gene onto the short arm of chromosome 8, using 21 affected Japanese families. The gene is located between marker D8S87 and ankyrin (ANK1). More recently, more refined mapping has pinpointed the WS gene to a region between marker D8S131 and D8S87, an 8.3 cM interval. Identification of the gene and gene product should add considerably to understanding the basis of Werner Syndrome and enable biochemical and genetic approaches to diagnosis and treatment.

The present invention provides a novel, previously unidentified gene for Werner Syndrome and compositions for diagnosis and treatment of WS, and further provides other related advantages.

SUMMARY OF THE INVENTION

Briefly stated, the present invention provides isolated nucleic acid molecules encoding the WRN gene, as well as portions thereof, representative of which are provided in the Figures. The protein which is encoded by the WRN gene is referred to hereinafter as the "WRN protein". Within other embodiments, nucleic acid molecules are provided which encode a mutant WRN gene product that increases the

Within other aspects of the present invention, isolated nucleic acid molecules are provided, selected from the group consisting of (a) an isolated nucleic acid molecule as set forth in the Figures, or complementary sequence thereof, (b) an isolated nucleic acid molecule that specifically hybridizes to the nucleic acid molecule of (a) under conditions of high stringency, and (c) an isolated nucleic acid that encodes a WRN gene product (WRN protein). As utilized herein, it should be understood that a nucleic acid molecule hybridizes "specifically" to an WRN gene (or related sequence) if it hybridizes detectably to such a sequence, but does not significantly or detectably hybridize to the Bloom's Syndrome gene (Ellis et al., *Cell* 83:655-666, 1995).

Within other aspects, expression vectors are provided comprising a promoter operably linked to one of the nucleic acid molecule described above. Representative examples of suitable promoters include tissue-specific promoters, as well as promoters such as the CMV I-E promoter, SV40 early promoter and Mol.V LTR. Within related aspects, viral vectors are provided that are capable of directing the expression of a nucleic acid molecule as described above. Representative examples of such viral vectors include herpes simplex viral vectors, adenoviral vectors, adenovirus-associated viral vectors and retroviral vectors. Also provided are host cells (e.g., human, dog, monkey, rat or mouse cells) which carry the above-described vectors.

Within other aspects of the present invention, isolated proteins or polypeptides are provided comprising a WRN gene product, as well as peptides of greater than 12, 13 or 20 amino acids. Within another embodiment, the protein is a mutant WRN gene product that increases the probability of Werner's Syndrome.

Within yet another aspect of the present invention, methods of treating or preventing Werner's Syndrome are provided (as well as for related diseases which are discussed in more detail below), comprising the step of administering to a patient a vector containing or expressing a nucleic acid molecule as described above, thereby reducing the likelihood or delaying the onset of Werner's Syndrome (or the related

Figures 1-4 are schematic diagrams of

Figures 5-8 are schematic diagrams of

administering to a patient a protein as described above, thereby reducing the likelihood

or delaying the onset of Werner's Syndrome (or a related disease) in the patient. Within certain embodiments, the above methods may be accomplished by *in vivo* administration.

Also provided by the present invention are pharmaceutical compositions comprising a nucleic acid molecule, vector, host cell, protein, or antibody as described above, along with a pharmaceutically acceptable carrier or diluent.

Within other aspects of the present invention, antibodies are provided which specifically bind to an WRN protein or to unique peptides derived therefrom. As utilized herein, it should be understood that an antibody is specific for an WRN protein (or peptide) if it binds detectably, and with a K_D of $10^{-7}M$ or less (e.g., $10^{-8}M$, $10^{-9}M$, etc.), but does not bind detectably (or with an affinity of greater than $10^{-7}M$, (e.g., $10^{-6}M$, $10^{-5}M$, etc.) to an unrelated helicase (e.g., the Bloom's Syndrome gene, *supra*). Also provided are hybridomas which are capable of producing such antibodies.

Within other aspects of the present invention, nucleic acid probes are provided which are capable of specifically hybridizing (as defined below) to an WRN gene under conditions of high stringency. Within one related aspect, such probes comprise at least a portion of the nucleotide sequence shown in the Figures, or its complementary sequence, the probe being capable of specifically hybridizing to a mutant WRN gene under conditions of high stringency. Representative probes of the present invention are generally at least 12 nucleotide bases in length, although they may be 14, 16, 18 bases or longer. Also provided are primer pairs capable of specifically amplifying all or a portion of any of the nucleic acid molecules disclosed herein.

Within other aspects of the invention, methods are provided for diagnosing a patient having an increased likelihood of contracting Werner's Syndrome (or a related disease), comprising the steps of (a) obtaining from a patient a biological sample containing nucleic acid, (b) incubating the nucleic acid with a probe which is capable of specifically hybridizing to a mutant WRN gene under conditions of high stringency, and (c) determining that said patient has an increased likelihood of contracting Werner's Syndrome (or a related disease). Within another aspect,

the method further comprises determining that said patient has an increased likelihood of contracting Werner's Syndrome (or a related disease). Within another aspect,

methods are provided comprising the steps of (a) obtaining from a patient a biological sample containing nucleic acid, (b) amplifying a selected nucleic acid sequence associated with a mutant WRN gene, and (c) detecting the presence of an amplified nucleic acid sequence, and thereby determining that the patient has an increased
5 likelihood of contracting Werner's Syndrome (or a related disease). Suitable biological samples include nucleated cells obtained from the peripheral blood, from buccal swabs, or brain tissue.

Within another aspect, peptide vaccines are provided which comprise a portion of a mutant WRN gene product containing a mutation, in combination with a
10 pharmaceutically acceptable carrier or diluent.

Within yet another aspect, transgenic animals are provided whose germ cells and somatic cells contain a WRN gene (or lack thereof, *i.e.*, a "knockout") which is operably linked to a promoter effective for the expression of the gene, the gene being introduced into the animal, or an ancestor of the animal, at an embryonic stage. Within
15 one embodiment, the animal is a mouse, rat or dog. Within other embodiments, the WRN gene is expressed from a vector as described above. Within yet another embodiment, the WRN gene encodes a mutant WRN gene product.

These and other aspects of the present invention will become evident upon reference to the following detailed description and attached drawings. In addition,
20 various references are set forth herein which describe in more detail certain procedures or compositions (*e.g.*, plasmids, etc.), and are therefore incorporated by reference in their entirety.

BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCE LISTING

25 Figure 1 is a genetic and physical map of the WRN region. The genetic map (A) of the region is sex-equal with distances given in cM. The polymorphic loci used (B) are di-nucleotide and tri-nucleotide repeat STRP loci. The physical map

30 2253, 3833, 2236, and 3101. Marker order was determined from the sequence-tagged

site (STS) content of YACs, P1 clones, and cosmid clones and from genomic DNA sequence from P1 clones. The YACs presented (D) represent the minimal tiling and are the YACs used for cDNA selection experiments. The P1 and cosmid clones needed for the minimum tiling path are shown (E). Clones shown are P1 clones except for 8C11, which is a cosmid clone. Clone order was established by STS content.

Figures 2A and 2B are the DNA (SEQ ID No. 70) and predicted amino acid (SEQ ID No. 71) sequences of the WRN gene transcript. The one-letter amino acid code is used in Fig. 2B.

Figures 3A-3C are the DNA and predicted amino acid sequence of an alternate WRN gene transcript (SEQ ID Nos. 72 and 73).

Figures 4A-4G are an alignment of the WRN gene product (SEQ ID No. 74) with known helicases from *S. pombe* (SEQ ID No. 76), *E. coli* (SEQ ID No. 75), human (SEQ ID No. 77) and the Bloom's Syndrome gene "BLM" (SEQ ID No. 78).

Figures 5A-5U are the genomic DNA sequence of the region containing a WRN gene (SEQ ID No. 79).

Figure 6 presents a cDNA sequence of the mouse WRN gene (SEQ ID Nos. 205 and 206).

Figure 7 is a genomic DNA sequence of the mouse WRN gene (SEQ ID Nos. 207-209).

Figure 8 is a diagram of the WRN gene product with location of mutations. A, WRN cDNA. Numbering across the top refers to the cDNA sequence as numbered in GenBank L76937. B, Predicted WRN gene product. The helicase domain is designated as "HD", motifs from I to VI are indicated. C, Location of mutations. Numbering across the bottom refer to the mutations. *: nonsense mutation. ^: frame shift mutation caused by a single base deletion. Gray lines: frame shift mutations causing deletion of exon(s). D, Predicted proteins. Lines represent the different predicted truncated proteins produced from mutations in the WRN gene.

Figure 9 shows WRN gene product by indirect antibody staining (panel A), Western (panel B), and the size of cells (panel C) expressing the WRN gene.

Figure 10 shows the alignment of the mouse and human WRN gene products.

DETAILED DESCRIPTION OF THE INVENTION

5 Definitions

Prior to setting forth the invention in detail, it may be helpful to an understanding thereof to set forth definitions of certain terms and to list and to define the abbreviations that will be used hereinafter.

10 "Genetic marker" is any segment of a chromosome that is distinguishably unique in the genome, and polymorphic in the population so as to provide information about the inheritance of linked DNA sequences, genes and/or other markers.

 "Vector" refers to an assembly which is capable of directing the expression of a WRN gene, as well as any additional sequence(s) or gene(s) of interest.
15 The vector must include transcriptional promoter elements which are operably linked to the genes of interest. The vector may be composed of either deoxyribonucleic acids ("DNA"), ribonucleic acids ("RNA"), or a combination of the two (e.g., a DNA-RNA chimeric). Optionally, the vector may include a polyadenylation sequence, one or more restriction sites, as well as one or more selectable markers such as neomycin
20 phosphotransferase or hygromycin phosphotransferase. Additionally, depending on the host cell chosen and the vector employed, other genetic elements such as an origin of replication, additional nucleic acid restriction sites, enhancers, sequences conferring inducibility of transcription, and selectable markers, may also be incorporated into the vectors described herein.

25 Abbreviations: YAC, yeast artificial chromosome; EST, expressed sequence tag; PCR, polymerase chain reaction; RT-PCR, PCR process in which RNA is first transcribed into DNA at the first step using reverse transcriptase (RT); cDNA, any

30 compositions for the detection and treatment of Werner's Syndrome, as well as related

diseases. These methods and compositions include a family of Werner's Syndrome-related genes, and the proteins encoded thereby, that have been implicated in the onset of Werner's Syndrome. These genes and proteins, including genetic markers, nucleic acid sequences and clones, are also useful in the creation of *in vitro* and animal models and screening tests useful for the study of Werner's Syndrome, including the possible identification of other genes implicated in Werner's Syndrome. The present invention also provides vector constructs, genetic markers, nucleic acid sequences, clones, diagnostic tests and compositions and methods for the identification of individuals likely to suffer from Werner's Syndrome.

Genes and Gene Products Related To Werner's Syndrome

The present invention provides isolated nucleic acid molecules comprising a portion of the gene which is implicated in the onset of WS. Briefly, as can be seen from Figure 4, this gene encodes a protein that is similar in amino acid sequence to several known ATP-dependent DNA helicases (enzymes that unwind the DNA duplex). It is less similar to known RNA-DNA helicases. Helicases are involved in the replication of DNA, often binding the replication origin, and/or the replication complex. In addition, the single stranded DNA that is involved in recombination can be generated by DNA helicases.

Although various aspects of the WRN gene (or portions thereof) are shown in the Figures, it should be understood that within the context of the present invention, reference to one or more of these genes includes derivatives of the genes that are substantially similar to the genes (and, where appropriate, the proteins (including peptides and polypeptides) that are encoded by the genes and their derivatives). As used herein, a nucleotide sequence is deemed to be "substantially similar" if: (a) the nucleotide sequence is derived from the coding region of the described genes and includes, for example, portions of the sequence or allelic variations of the sequences

hybridization to nucleotide sequences of the present invention under high or very high

stringency (see Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, NY, 1989); or (c) the nucleic acid sequences are degenerate as a result of the genetic code to the nucleic acid sequences defined in (a) or (b). Further, the nucleic acid molecule disclosed herein includes both complementary and non-complementary sequences, provided the sequences otherwise meet the criteria set forth herein. Within the context of the present invention, high stringency means standard hybridization conditions (e.g., 5x SSPE, 0.5% SDS at 65°C, or the equivalent) while very high stringency means conditions of hybridization such that the nucleotide sequence is able to selectively hybridize to a single allele of the WRN-related gene.

10 The WRN gene may be isolated from genomic DNA or cDNA. Genomic DNA libraries constructed in chromosomal vectors, such as YACs (yeast artificial chromosomes), bacteriophage vectors, such as λ EMBL3, λ gt10, cosmids, or plasmids are suitable for use. cDNA libraries constructed in bacteriophage vectors, plasmids, or others, are suitable for screening. Such libraries may be constructed using
15 methods and techniques known in the art (see Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press, 1989) or purchased from commercial sources (e.g., Clontech, Palo Alto, CA). Within one embodiment, the WRN gene is isolated by PCR performed on genomic DNA, cDNA or DNA from libraries, or is isolated by probe hybridization of genomic DNA or cDNA libraries. Primers for PCR
20 and probes for hybridization screening may be designed based on the DNA sequence of WRN presented herein. The DNA sequence of a portion of the WRN gene and the entire coding sequence is presented in the Figures. Primers for PCR should be derived from sequences in the 5' and 3' untranslated region in order to isolate a full-length cDNA. The primers should not have self-complementary sequences nor have
25 complementary sequences at their 3' end (to prevent primer-dimer formation). Preferably, the primers have a GC content of about 50% and contain restriction sites. The primers are annealed to cDNA and sufficient cycles of PCR are performed to yield

propagated. An oligonucleotide hybridization probe suitable for screening genomic or

cDNA libraries may be designed based on the sequence provided herein. Preferably, the oligonucleotide is 20-30 bases long. Such an oligonucleotide may be synthesized by automated synthesis. The oligonucleotide may be conveniently labeled at the 5' end with a reporter molecule, such as a radionuclide, (e.g., ^{32}P) or biotin. The library is
5 plated as colonies or phage, depending upon the vector, and the recombinant DNA is transferred to nylon or nitrocellulose membranes. Following denaturation, neutralization, and fixation of the DNA to the membrane, the membranes are hybridized with the labeled probe. The membranes are washed and the reporter molecule detected. The hybridizing colonies or phage are isolated and propagated. Candidate clones or
10 PCR amplified fragments may be verified as containing WRN DNA by any of various means. For example, the candidate clones may be hybridized with a second, nonoverlapping probe or subjected to DNA sequence analysis. In these ways, clones containing WRN gene, which are suitable for use in the present invention are isolated.

The structure of the proteins encoded by the nucleic acid molecules
15 described herein may be predicted from the primary translation products using the hydrophobicity plot function of, for example, P/C Gene, Lasergen System, DNA STAR, Madison, Wisconsin, or according to the methods described by Kyte and Doolittle (*J. Mol. Biol.* 157:105-132, 1982).

WRN proteins of the present invention may be prepared in the form of
20 acidic or basic salts, or in neutral form. In addition, individual amino acid residues may be modified by oxidation or reduction. Furthermore, various substitutions, deletions, or additions may be made to the amino acid or nucleic acid sequences, the net effect of which is to retain or further enhance or decrease the biological activity of the mutant or wild-type protein. Moreover, due to degeneracy in the genetic code, for example, there
25 may be considerable variation in nucleotide sequences encoding the same amino acid sequence.

Other derivatives of the WRN proteins disclosed herein include

which may be added to facilitate purification or identification of WRN proteins (see

U.S. Patent No. 4,851,341; *see also*, Hopp et al., *Bio/Technology* 6:1204, 1988.) Alternatively, fusion proteins such as WRN protein- β -galactosidase or WRN protein-luciferase may be constructed in order to assist in the identification, expression, and analysis of WRN proteins.

5 WRN proteins of the present invention may be constructed using a wide variety of techniques described herein. Further, mutations may be introduced at particular loci by synthesizing oligonucleotides containing a mutant sequence, flanked by restriction sites enabling ligation to fragments of the native sequence. Following ligation, the resulting reconstructed sequence encodes a derivative having the desired
10 amino acid insertion, substitution, or deletion.

Alternatively, oligonucleotide-directed site-specific (or segment specific) mutagenesis procedures may be employed to provide an altered gene having particular codons altered according to the substitution, deletion, or insertion required. Exemplary methods of making the alterations set forth above are disclosed by Walder et al. (*Gene*
15 42:133, 1986); Bauer et al. (*Gene* 37:73, 1985); Craik (*BioTechniques*, January 1985, 12-19); Smith et al. (*Genetic Engineering: Principles and Methods*, Plenum Press, 1981); and Sambrook et al. (*supra*). Deletion or truncation derivatives of WRN proteins (*e.g.*, a soluble extracellular portion) may also be constructed by utilizing convenient restriction endonuclease sites adjacent to the desired deletion. Subsequent
20 to restriction, overhangs may be filled in, and the DNA religated. Exemplary methods of making the alterations set forth above are disclosed by Sambrook et al. (*Molecular Cloning: A Laboratory Manual*, 2d Ed., Cold Spring Harbor Laboratory Press, 1989).

Mutations of the present invention preferably preserve the reading frame of the coding sequences. Furthermore, the mutations will preferably not create
25 complementary regions that could hybridize to produce secondary mRNA structures, such as loops or hairpins, that would adversely affect translation of the mRNA. Although a mutation site may be predetermined, it is not necessary that the nature of the

target codon and the expressed mutants screened for indicative biological activity

Alternatively, mutations may be introduced at particular loci by synthesizing oligonucleotides containing a mutant sequence, flanked by restriction sites enabling ligation to fragments of the native sequence. Following ligation, the resulting reconstructed sequence encodes a derivative having the desired amino acid insertion, substitution, or deletion.

WRN proteins may also be constructed utilizing techniques of PCR mutagenesis, chemical mutagenesis (Drinkwater and Klinedinst, *PNAS* 83:3402-3406, 1986), by forced nucleotide misincorporation (e.g., Liao and Wise *Gene* 88:107-111, 1990), or by use of randomly mutagenized oligonucleotides (Horwitz et al., *Genome* 3:112-117, 1989).

Proteins can be isolated by, among other methods, culturing suitable host and vector systems to produce the recombinant translation products of the present invention. Supernates from such cell lines, or protein inclusions or whole cells where the protein is not excreted into the supernate, can then be treated by a variety of purification procedures in order to isolate the desired proteins. For example, the supernate may be first concentrated using commercially available protein concentration filters, such as an Amicon or Millipore Pellicon ultrafiltration unit. Following concentration, the concentrate may be applied to a suitable purification matrix such as, for example, an anti-protein antibody bound to a suitable support. Alternatively, anion or cation exchange resins may be employed in order to purify the protein. As a further alternative, one or more reverse-phase high performance liquid chromatography (RP-HPLC) steps may be employed to further purify the protein. Other methods of isolating the proteins of the present invention are well known in the skill of the art.

A protein is deemed to be "isolated" within the context of the present invention if no other (undesired) protein is detected pursuant to sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis followed by Coomassie blue staining. Within other embodiments, the desired protein can be isolated such that no

Expression of a WRN gene

The present invention also provides for the manipulation and expression of the above described genes by culturing host cells containing a vector capable of expressing the above-described genes. Such vectors or vector constructs include either
5 synthetic or cDNA-derived nucleic acid molecules encoding WRN proteins, which are operably linked to suitable transcriptional or translational regulatory elements. Suitable regulatory elements may be derived from a variety of sources, including bacterial, fungal, viral, mammalian, insect, or plant genes. Selection of appropriate regulatory elements is dependent on the host cell chosen, and may be readily accomplished by one
10 of ordinary skill in the art. Examples of regulatory elements include: a transcriptional promoter and enhancer or RNA polymerase binding sequence, a transcriptional terminator, and a ribosomal binding sequence, including a translation initiation signal.

Nucleic acid molecules that encode any of the WRN proteins described above may be readily expressed by a wide variety of prokaryotic and eukaryotic host
15 cells, including bacterial, mammalian, yeast or other fungi, viral, insect, or plant cells. Methods for transforming or transfecting such cells to express foreign DNA are well known in the art (see, e.g., Itakura et al., U.S. Patent No. 4,704,362; Hinnen et al., *Proc. Natl. Acad. Sci. USA* 75:1929-1933, 1978; Murray et al., U.S. Patent No. 4,801,542; Upshall et al., U.S. Patent No. 4,935,349; Hagen et al., U.S. Patent No. 4,784,950; Axel
20 et al., U.S. Patent No. 4,399,216; Goeddel et al., U.S. Patent No. 4,766,075; and Sambrook et al. *Molecular Cloning: A Laboratory Manual*, 2nd edition, Cold Spring Harbor Laboratory Press, 1989; for plant cells see Czako and Marton, *Plant Physiol* 104:1067-1071, 1994; and Paszkowski et al., *Biotech.* 24:387-392, 1992).

Bacterial host cells suitable for carrying out the present invention include
25 *E. coli*, *B. subtilis*, *Salmonella typhimurium*, and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*, as well as many other bacterial species well known to one of ordinary skill in the art. Representative examples of functions in the host cell, one or more selectable phenotypic markers, and a bacterial

origin of replication. Representative promoters include the β -lactamase (penicillinase) and lactose promoter system (see Chang et al., *Nature* 275:615, 1978), the T7 RNA polymerase promoter (Studier et al., *Meth. Enzymol.* 185:60-89, 1990), the lambda promoter (Elvin et al., *Gene* 87:123-126, 1990), the *trp* promoter (Nichols and Yanofsky, *Meth. in Enzymology* 101:155, 1983) and the *tac* promoter (Russell et al., *Gene* 20: 231, 1982). Representative selectable markers include various antibiotic resistance markers such as the kanamycin or ampicillin resistance genes. Many plasmids suitable for transforming host cells are well known in the art, including among others, pBR322 (see Bolivar et al., *Gene* 2:95, 1977), the pUC plasmids pUC18, pUC19, pUC118, pUC119 (see Messing, *Meth. in Enzymology* 101:20-77, 1983 and Vieira and Messing, *Gene* 19:259-268, 1982), and pNH8A, pNH16a, pNH18a, and Bluescript M13 (Stratagene, La Jolla, Calif.).

Yeast and fungi host cells suitable for carrying out the present invention include, among others, *Saccharomyces pombe*, *Saccharomyces cerevisiae*, the genera *Pichia* or *Kluyveromyces* and various species of the genus *Aspergillus* (McKnight et al., U.S. Patent No. 4,935,349). Suitable expression vectors for yeast and fungi include, among others, YCp50 (ATCC No. 37419) for yeast, and the amdS cloning vector pV3 (Turnbull, *Bio/Technology* 7:169, 1989), YRp7 (Struhl et al., *Proc. Natl. Acad. Sci. USA* 76:1035-1039, 1978), YEp13 (Broach et al., *Gene* 8:121-133, 1979), pJDB249 and pJDB219 (Beggs, *Nature* 275:104-108, 1978) and derivatives thereof.

Preferred promoters for use in yeast include promoters from yeast glycolytic genes (Hitzeman et al., *J. Biol. Chem.* 255:12073-12080, 1980; Alber and Kawasaki, *J. Mol. Appl. Genet.* 1:419-434, 1982) or alcohol dehydrogenase genes (Young et al., in *Genetic Engineering of Microorganisms for Chemicals*, Hollaender et al. (eds.), p. 355, Plenum, New York, 1982; Ammerer, *Meth. Enzymol.* 101:192-201, 1983). Examples of useful promoters for fungi vectors include those derived from *Aspergillus nidulans* glycolytic genes, such as the *adh3* promoter (McKnight et al.

et al., *ibid.*, 1985).

As with bacterial vectors, the yeast vectors will generally include a selectable marker, which may be one of any number of genes that exhibit a dominant phenotype for which a phenotypic assay exists to enable transformants to be selected. Preferred selectable markers are those that complement host cell auxotrophy, provide
5 antibiotic resistance or enable a cell to utilize specific carbon sources, and include *leu2* (Broach et al., *ibid.*), *ura3* (Botstein et al., *Gene* 8:17, 1979), or *his3* (Struhl et al., *ibid.*). Another suitable selectable marker is the *cat* gene, which confers chloramphenicol resistance on yeast cells.

Techniques for transforming fungi are well known in the literature, and
10 have been described, for instance, by Beggs (*ibid.*), Hinnen et al. (*Proc. Natl. Acad. Sci. USA* 75:1929-1933, 1978), Yelton et al. (*Proc. Natl. Acad. Sci. USA* 81:1740-1747, 1984), and Russell (*Nature* 301:167-169, 1983). The genotype of the host cell may contain a genetic defect that is complemented by the selectable marker present on the expression vector. Choice of a particular host and selectable marker is well within the
15 level of ordinary skill in the art.

Protocols for the transformation of yeast are also well known to those of ordinary skill in the art. For example, transformation may be readily accomplished either by preparation of spheroplasts of yeast with DNA (*see* Hinnen et al., *PNAS USA* 75:1929, 1978) or by treatment with alkaline salts such as LiCl (*see* Itoh et al., *J*
20 *Bacteriology* 153:163, 1983). Transformation of fungi may also be carried out using polyethylene glycol as described by Cullen et al. (*Bio/Technology* 5:369, 1987).

Viral vectors include those which comprise a promoter that directs the expression of an isolated nucleic acid molecule that encodes an WRN protein as described above. A wide variety of promoters may be utilized within the context of the
25 present invention, including for example, promoters such as MoMLV LTR, RSV LTR, Friend MuLV LTR, adenoviral promoter (Ohno et al., *Science* 265: 781-784, 1994), neomycin phosphotransferase promoter/enhancer, late parvovirus promoter (Koering

et al., *Proc. Natl. Acad. Sci. USA* 84:1008-1012, 1987), and the cytomegalovirus immediate late promoter. Within particularly

preferred embodiments of the invention, the promoter is a tissue-specific promoter (see e.g., WO 91/02805; EP 0,415,731; and WO 90/07936). Representative examples of suitable tissue specific promoters include neural specific enolase promoter, platelet derived growth factor beta promoter, bone morpho-genetic protein promoter, human alpha1-chimaerin promoter, synapsin I promoter and synapsin II promoter. In addition to the above-noted promoters, other viral-specific promoters (e.g., retroviral promoters (including those noted above, as well as others such as HIV promoters), hepatitis, herpes (e.g., EBV), and bacterial, fungal or parasitic (e.g., malarial) -specific promoters may be utilized in order to target a specific cell or tissue which is infected with a virus, bacteria, fungus or parasite.

Thus, WRN proteins of the present invention may be expressed from a variety of viral vectors, including for example, herpes viral vectors (e.g., U.S. Patent No. 5,288,641), adenoviral vectors (e.g., WO 94/26914, WO 93/9191; Kolls et al., *PNAS* 91(1):215-219, 1994; Kass-Eisler et al., *PNAS* 90(24):11498-502, 1993; Guzman et al., *Circulation* 88(6):2838-48, 1993; Guzman et al., *Cir. Res* 73(6):1202-1207, 1993; Zabner et al., *Cell* 75(2):207-216, 1993; Li et al., *Hum Gene Ther.* 4(4):403-409, 1993; Caillaud et al., *Eur. J. Neurosci.* 5(10):1287-1291, 1993; Vincent et al., *Nat. Genet.* 5(2):130-134, 1993; Jaffe et al., *Nat. Genet.* 1(5):372-378, 1992; and Levrero et al., *Gene* 101(2):195-202, 1991), adeno-associated viral vectors (WO 95/13365; Flotte et al., *PNAS* 90(22):10613-10617, 1993), baculovirus vectors, parvovirus vectors (Koering et al., *Hum. Gene Therap.* 5:457-463, 1994), pox virus vectors (Panicali and Paoletti, *PNAS* 79:4927-4931, 1982; and Ozaki et al., *Biochem. Biophys. Res. Comm.* 193(2):653-660, 1993), and retroviruses (e.g., EP 0,415,731; WO 90/07936; WO 91/0285, WO 94/03622; WO 93/25698; WO 93/25234; U.S. Patent No. 5,219,740; WO 93/11230; WO 93/10218. Viral vectors may likewise be constructed which contain a mixture of different elements (e.g., promoters, envelope sequences and the like) from different viruses, or non-viral sources. Within various embodiments, either the viral

Mammalian cells suitable for carrying out the present invention include, among others: PC12 (ATCC No. CRL1721), N1E-115 neuroblastoma, SK-N-BE(2)C neuroblastoma, SHSY5 adrenergic neuroblastoma, NS20Y and NG108-15 murine cholinergic cell lines, or rat F2 dorsal root ganglion line, COS (*e.g.*, ATCC No. CRL 1650 or 1651), BHK (*e.g.*, ATCC No. CRL 6281; BHK 570 cell line (deposited with the American Type Culture Collection under accession number CRL 10314), CHO (ATCC No. CCL 61), HeLa (*e.g.*, ATCC No. CCL 2), 293 (ATCC No. 1573; Graham et al., *J. Gen. Virol.* 36:59-72, 1977) and NS-1 cells. Other mammalian cell lines may be used within the present invention, including Rat Hep I (ATCC No. CRL 1600), Rat Hep II (ATCC No. CRL 1548), TCMK (ATCC No. CCL 139), Human lung (ATCC No. CCL 75.1), Human hepatoma (ATCC No. HTB-52), Hep G2 (ATCC No. HB 8065), Mouse liver (ATCC No. CCL 29.1), NCTC 1469 (ATCC No. CCL 9.1), SP2/0-Ag14 (ATCC No. 1581), HIT-T15 (ATCC No. CRL 1777), and RINm 5AHT₂B (Orskov and Nielson, *FEBS* 229(1):175-178, 1988).

Mammalian expression vectors for use in carrying out the present invention will include a promoter capable of directing the transcription of a cloned gene or cDNA. Preferred promoters include viral promoters and cellular promoters. Viral promoters include the cytomegalovirus immediate early promoter (Boshart et al., *Cell* 41:521-530, 1985), cytomegalovirus immediate late promoter, SV40 promoter (Subramani et al., *Mol. Cell. Biol.* 1:854-864, 1981), MMTV LTR, RSV LTR, metallothionein-1, adenovirus E1a. Cellular promoters include the mouse metallothionein-1 promoter (Palmiter et al., U.S. Patent No. 4,579,821), a mouse V_K promoter (Bergman et al., *Proc. Natl. Acad. Sci. USA* 81:7041-7045, 1983; Grant et al., *Nucl. Acids Res.* 15:5496, 1987) and a mouse V_H promoter (Loh et al., *Cell* 33:85-93, 1983). The choice of promoter will depend, at least in part, upon the level of expression desired or the recipient cell line to be transfected.

Such expression vectors may also contain a set of RNA splice sites

adenovirus and/or immunoglobulin genes. Also contained in the expression vectors is a

polyadenylation signal located downstream of the coding sequence of interest. Suitable polyadenylation signals include the early or late polyadenylation signals from SV40 (Kaufman and Sharp, *ibid.*), the polyadenylation signal from the Adenovirus 5 E1B region and the human growth hormone gene terminator (DeNoto et al., *Nuc. Acids Res.* 9:3719-3730, 1981). The expression vectors may include a noncoding viral leader sequence, such as the Adenovirus 2 tripartite leader, located between the promoter and the RNA splice sites. Preferred vectors may also include enhancer sequences, such as the SV40 enhancer. Expression vectors may also include sequences encoding the adenovirus VA RNAs. Suitable expression vectors can be obtained from commercial sources (e.g., Stratagene, La Jolla, Calif.).

Vector constructs comprising cloned DNA sequences can be introduced into cultured mammalian cells by, for example, calcium phosphate-mediated transfection (Wigler et al., *Cell* 14:725, 1978; Corsaro and Pearson, *Somatic Cell Genetics* 7:603, 1981; Graham and Van der Eb, *Virology* 52:456, 1973), electroporation (Neumann et al., *EMBO J.* 1:841-845, 1982), or DEAE-dextran mediated transfection (Ausubel et al. (eds.), *Current Protocols in Molecular Biology*, John Wiley and Sons, Inc., NY, 1987). To identify cells that have stably integrated the cloned DNA, a selectable marker is generally introduced into the cells along with the gene or cDNA of interest. Preferred selectable markers for use in cultured mammalian cells include genes that confer resistance to drugs, such as neomycin, hygromycin, and methotrexate. The selectable marker may be an amplifiable selectable marker. Preferred amplifiable selectable markers are the DHFR gene and the neomycin resistance gene. Selectable markers are reviewed by Thilly (*Mammalian Cell Technology*, Butterworth Publishers, Stoneham, MA, which is incorporated herein by reference).

Mammalian cells containing a suitable vector are allowed to grow for a period of time, typically 1-2 days, to begin expressing the DNA sequence(s) of interest. Drug selection is then applied to select for growth of cells that are expressing the

amplifiable selectable marker. A drug concentration is then increased in a stepwise manner to select for increased copy number of the cloned sequences, thereby increasing

expression levels. Cells expressing the introduced sequences are selected and screened for production of the protein of interest in the desired form or at the desired level. Cells that satisfy these criteria can then be cloned and scaled up for production.

5 Protocols for the transfection of mammalian cells are well known to those of ordinary skill in the art. Representative methods include calcium phosphate mediated transfection, electroporation, lipofection, retroviral, adenoviral and protoplast fusion-mediated transfection (*see* Sambrook et al., *supra*). Naked vector constructs can also be taken up by muscular cells or other suitable cells subsequent to injection into the muscle of a mammal (or other animals).

10 Numerous insect host cells known in the art can also be useful within the present invention, in light of the subject specification. For example, the use of baculoviruses as vectors for expressing heterologous DNA sequences in insect cells has been reviewed by Atkinson et al. (*Pestic. Sci.* 28:215-224,1990).

15 Numerous plant host cells known in the art can also be useful within the present invention, in light of the subject specification. For example, the use of *Agrobacterium rhizogenes* as vectors for expressing genes in plant cells has been reviewed by Sinkar et al., (*J. Biosci. (Bangalore)* 11:47-58, 1987).

20 WRN proteins may be prepared by growing (typically by culturing) the host/vector systems described above, in order to express the recombinant WRN proteins. Recombinantly produced WRN proteins may be further purified as described in more detail below.

25 Within related aspects of the present invention, WRN proteins may be expressed in a transgenic animal whose germ cells and somatic cells contain a WRN gene which is operably linked to a promoter effective for the expression of the gene. Alternatively, in a similar manner transgenic animals may be prepared that lack the WRN gene (*e.g.*, "knockout" mice). Such transgenics may be prepared in a variety non-human animals, including mice, rats, rabbits, sheep, dogs, goats and pigs (*see* Hammer

41:343-345, 1985 and U.S. Patent Nos. 5,175,383, 5,087,571, 4,736,866, 5,387,742, 5,347,075, 5,221,778, and 5,175,384).

Briefly, an expression vector, including a nucleic acid molecule to be expressed together with appropriately positioned expression control sequences, is introduced into pronuclei of fertilized eggs, for example, by microinjection. Integration of the injected DNA is detected by blot analysis of DNA from tissue samples. It is preferred that the introduced DNA be incorporated into the germ line of the animal so that it is passed on to the animal's progeny. Tissue-specific expression may be achieved through the use of a tissue-specific promoter, or through the use of an inducible promoter, such as the metallothionein gene promoter (Palmiter et al., 1983, *ibid*), which allows regulated expression of the transgene.

Vectors of the present invention may contain or express a wide variety of additional nucleic acid molecules in place of or in addition to an WRN protein as described above, either from one or several separate promoters. For example, the viral vector may express a lymphokine or lymphokine receptor, antisense or ribozyme sequence or toxins. Representative examples of lymphokines include IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, GM-CSF, G-CSF, M-CSF, alpha-interferon, beta-interferon, gamma-interferon, and tumor necrosis factors, as well as their respective receptors. Representative examples of antisense sequences include antisense sequences which block the expression of WRN protein mutants. Representative examples of toxins include: ricin, abrin, diphtheria toxin, cholera toxin, saporin, gelonin, pokeweed antiviral protein, tritin, *Shigella* toxin, and *Pseudomonas* exotoxin A.

Within other aspects of the invention, antisense oligonucleotide molecules are provided which specifically inhibit expression of mutant WRN nucleic acid sequences (*see generally*, Hirashima et al. in *Molecular Biology of RNA: New Perspectives* (M. Inouye and B. S. Dudock, eds., 1987 Academic Press, San Diego, p.

199-206). U.S. 5,359,951; WO 92/06693; and EP-A2-612844). Briefly, such

molecules are constructed such that they are complementary to, and able to form Watson-Crick base pairs with, a region of transcribed WRN mutant mRNA sequence containing an WRN mutation. The resultant double-stranded nucleic acid interferes with subsequent processing of the mRNA, thereby preventing protein synthesis.

5 Within other related aspects of the invention, ribozyme molecules are provided wherein an antisense oligonucleotide sequence is incorporated into a ribozyme which can specifically cleave mRNA molecules transcribed from a mutant WRN gene (see generally, Kim et al. *Proc. Nat. Acad. Sci. USA* 84:8788 (1987); Haseloff, et al. *Nature* 234:585 (1988), Cech, *JAMA* 260:3030 (1988); Jeffries, et al. *Nucleic Acids Res.* 17:1371 (1989); U.S. 5,093,246; U.S. 5,354,855, U.S. 5,144,019; U.S. 5,272,262; U.S. 5,254,678; and U.S. 4,987,071). According to this aspect of the invention, the antisense
10 sequence which is incorporated into a ribozyme includes a sequence complementary to, and able to form Watson-Crick base pairs with, a region of the transcribed mutant WRN mRNA containing an WRN mutation. The antisense sequence thus becomes a targeting
15 agent for delivery of catalytic ribozyme activity specifically to mutant WRN mRNA, where such catalytic activity cleaves the mRNA to render it incapable of being subsequently processed for WRN protein translation.

Host Cells

20 As discussed above, nucleic acid molecules which encode the WRN proteins of the present invention (or the vectors which contain and/or express related mutants) may readily be introduced into a wide variety of host cells. Representative examples of such host cells include plant cells, eukaryotic cells, and prokaryotic cells. Within preferred embodiments, the nucleic acid molecules are introduced into cells
25 from a vertebrate or warm-blooded animal, such as a human, macaque, dog, cow, horse, pig, sheep, rat, hamster, mouse or fish cell, or any hybrid thereof.

Preferred prokaryotic host cells for use within the present invention

include *Escherichia coli*.

Sequences cloned therein are well known in the art (see, e.g., Maniatis et al., *Molecular*

Cloning A Laboratory Manual, Cold Spring Harbor Laboratory, 1982, which is incorporated herein by reference; or Sambrook et al., *supra*). Vectors used for expressing cloned DNA sequences in bacterial hosts will generally contain a selectable marker, such as a gene for antibiotic resistance, and a promoter that functions in the host cell. Appropriate promoters include the *trp* (Nichols and Yanofsky, *Meth. Enzymol.* 101:155-164, 1983), *lac* (Casadaban et al., *J. Bacteriol.* 143:971-980, 1980), and phage λ (Queen, *J. Mol. Appl. Genet.* 2:1-10, 1983) promoter systems. Plasmids useful for transforming bacteria include the pUC plasmids (Messing, *Meth. Enzymol.* 101:20-78, 1983; Vieira and Messing, *Gene* 19:259-268, 1982), pBR322 (Bolivar et al., *Gene* 2:95-113, 1977), pCQV2 (Queen, *ibid.*), and derivatives thereof. Plasmids may contain both viral and bacterial elements.

Preferred eukaryotic cells include cultured mammalian cell lines (e.g., rodent or human cell lines) and fungal cells, including species of yeast (e.g., *Saccharomyces* spp., particularly *S. cerevisiae*, *Schizosaccharomyces* spp., or *Kluyveromyces* spp.) or filamentous fungi (e.g., *Aspergillus* spp., *Neurospora* spp.). Strains of the yeast *Saccharomyces cerevisiae* are particularly preferred. Methods for producing recombinant proteins in a variety of prokaryotic and eukaryotic host cells are generally known in the art (see, "Gene Expression Technology," *Methods in Enzymology*, Vol. 185, Goeddel (ed.), Academic Press, San Diego, Calif., 1990; see also, "Guide to Yeast Genetics and Molecular Biology," *Methods in Enzymology*, Guthrie and Fink (eds.), Academic Press, San Diego, Calif., 1991). In general, a host cell will be selected on the basis of its ability to produce the protein of interest at a high level or its ability to carry out at least some of the processing steps necessary for the biological activity of the protein. In this way, the number of cloned DNA sequences that must be introduced into the host cell can be minimized and overall yield of biologically active protein can be maximized.

The nucleic acid molecules (or vectors) may be introduced into host cells

Pearson, *Somatic Cell Genet.* 7:603, 1981, Graham and Van der Eb, *Virology* 52:456,

1973), electroporation (Neumann et al., *EMBO J.* 1:841-845, 1982), retroviral, adenoviral, protoplast fusion-mediated transfection or DEAE-dextran mediated transfection (Ausubel et al., (eds.), *Current Protocols in Molecular Biology*, John Wiley and Sons, Inc., NY, NY, 1987).

5 Host cells containing vector constructs of the present invention are then cultured to express a DNA molecule as described above. The cells are cultured according to standard methods in a culture medium containing nutrients required for growth of the chosen host cells. A variety of suitable media are known in the art and generally include a carbon source, a nitrogen source, essential amino acids, vitamins
10 and minerals, as well as other components, e.g., growth factors or serum, that may be required by the particular host cells. The growth medium will generally select for cells containing the DNA construct(s) by, for example, drug selection or deficiency in an essential nutrient which is complemented by the selectable marker on the DNA construct or co-transfected with the DNA construct.

15 Suitable growth conditions for yeast cells, for example, include culturing in a chemically defined medium, comprising a nitrogen source, which may be a non-amino acid nitrogen source or a yeast extract, inorganic salts, vitamins and essential amino acid supplements at a temperature between 4°C and 37°C, with 30°C being particularly preferred. The pH of the medium is preferably maintained at a pH greater
20 than 2 and less than 8, more preferably pH 5-6. Methods for maintaining a stable pH include buffering and constant pH control. Preferred agents for pH control include sodium hydroxide. Preferred buffering agents include succinic acid and Bis-Tris (Sigma Chemical Co., St. Louis, Mo.). Due to the tendency of yeast host cells to hyperglycosylate heterologous proteins, it may be preferable to express the nucleic acid
25 molecules of the present invention in yeast cells having a defect in a gene required for asparagine-linked glycosylation. Such cells are preferably grown in a medium containing an osmotic stabilizer. A preferred osmotic stabilizer is sorbitol

Cultured mammalian cells are generally cultured in commercially available serum-containing or serum-free media. Selection of a medium and growth conditions appropriate for the particular cell line used is well within the level of ordinary skill in the art.

5

Antibodies

Antibodies to the WRN proteins discussed above may readily be prepared given the disclosure provided herein. Such antibodies may, within certain embodiments, specifically recognize wild type WRN protein rather than a mutant WRN protein, mutant WRN protein rather than wild type WRN protein, or equally recognize both the mutant and wild-type forms of WRN protein. Antibodies may be used for isolation of the protein, establishing intracellular localization of the WRN protein, inhibiting activity of the protein (antagonist), or enhancing activity of the protein (agonist). Knowledge of the intracellular location of the WRN gene product may be abnormal in patients with WRN mutations, thus allowing the development of a rapid screening assay. As well, assays for small molecules that interact with the WRN gene product will be facilitated by the development of antibodies and localization studies.

Within the context of the present invention, antibodies are understood to include monoclonal antibodies, polyclonal antibodies, anti-idiotypic antibodies, antibody fragments (e.g., Fab, and F(ab')₂, F_v variable regions, or complementarity determining regions). As discussed above, antibodies are understood to be specific against an WRN protein if it binds with a K_d of greater than or equal to 10⁻⁷M, preferably greater than or equal to 10⁻⁸M. The affinity of a monoclonal antibody or binding partner can be readily determined by one of ordinary skill in the art (see Scatchard, *Ann. N.Y. Acad. Sci.* 51:660-672, 1949).

Briefly, polyclonal antibodies may be readily generated by one of ordinary skill in the art from a variety of warm-blooded animals such as horses, cows,

herein, and amino acids, preferably conjugated to a protein, either by chemical cross-linking (with glutaraldehyde) or utilized to immunize the animal, through

30

intraperitoneal, intramuscular, intraocular, or subcutaneous injections, an adjuvant such as Freund's complete or incomplete adjuvant. Merely as an example, a peptide corresponding to residues 1375 through 1387 of the WRN polypeptide sequence is used to raise a rabbit polyclonal antiserum. Following several booster immunizations, samples of serum are collected and tested for reactivity to the WRN protein or peptide. Particularly preferred polyclonal antisera will give a signal on one of these assays that is at least three times greater than background. Once the titer of the animal has reached a plateau in terms of its reactivity to the protein, larger quantities of antisera may be readily obtained either by weekly bleedings, or by exsanguinating the animal.

Monoclonal antibodies may also be readily generated using conventional techniques (see U.S. Patent Nos. RE 32,011, 4,902,614, 4,543,439, and 4,411,993 which are incorporated herein by reference; see also *Monoclonal Antibodies, Hybridomas: A New Dimension in Biological Analyses*, Plenum Press, Kennett, McKearn, and Bechtol (eds.), 1980, and *Antibodies: A Laboratory Manual*, Harlow and Lane (eds.), Cold Spring Harbor Laboratory Press, 1988, which are also incorporated herein by reference).

Briefly, within one embodiment a subject animal such as a rat or mouse is injected with an WRN protein or portion thereof as described above. The protein may be admixed with an adjuvant such as Freund's complete or incomplete adjuvant in order to increase the resultant immune response. Between one and three weeks after the initial immunization the animal may be reimmunized with another booster immunization, and tested for reactivity to the protein utilizing assays described above. Once the animal has reached a plateau in its reactivity to the injected protein, it is sacrificed, and organs which contain large numbers of B cells such as the spleen and lymph nodes are harvested.

Cells which are obtained from the immunized animal may be immortalized by transfection with a virus such as the Epstein-Barr virus (EBV) (see

suitable myeloma cell in order to create a 'hybridoma' which secretes monoclonal

antibody. Suitable myeloma lines include, for example, NS-1 (ATCC No. TIB 18), and P3X63 - Ag 8.653 (ATCC No. CRL 1580).

Following the fusion, the cells may be placed into culture plates containing a suitable medium, such as RPMI 1640, or DMEM (Dulbecco's Modified Eagles Medium) (JRH Biosciences, Lenexa, Kansas), as well as additional ingredients, such as fetal bovine serum (FBS, *i.e.*, from Hyclone, Logan, Utah, or JRH Biosciences). Additionally, the medium should contain a reagent which selectively allows for the growth of fused spleen and myeloma cells such as HAT (hypoxanthine, aminopterin, and thymidine) (Sigma Chemical Co., St. Louis, Missouri). After about seven days, the resulting fused cells or hybridomas may be screened in order to determine the presence of antibodies which are reactive against an WRN protein. A wide variety of assays may be utilized to determine the presence of antibodies which are reactive against the proteins of the present invention, including for example countercurrent immuno-electrophoresis, radioimmunoassays, radioimmunoprecipitations, enzyme-linked immuno-sorbent assays (ELISA), dot blot assays, western blots, immunoprecipitation, Inhibition or Competition Assays, and sandwich assays (*see* U.S. Patent Nos. 4,376,110 and 4,486,530; *see also* *Antibodies: A Laboratory Manual*, Harlow and Lane (eds.), Cold Spring Harbor Laboratory Press, 1988). Following several clonal dilutions and reassays, a hybridoma producing antibodies reactive against the WRN protein may be isolated.

Other techniques may also be utilized to construct monoclonal antibodies (*see* William D. Huse et al., "Generation of a Large Combinational Library of the Immunoglobulin Repertoire in Phage Lambda," *Science* 246:1275-1281, December 1989; *see also* L. Sastry et al., "Cloning of the Immunological Repertoire in *Escherichia coli* for Generation of Monoclonal Catalytic Antibodies: Construction of a Heavy Chain Variable Region-Specific cDNA Library," *Proc. Natl. Acad. Sci. USA* 86:5728-5732, August 1989; *see also* Michelle Alting-Mees et al., "Monoclonal Antibody Expression

La Jolla, California, which enables the production of antibodies through recombinant

techniques). Briefly, mRNA is isolated from a B cell population, and utilized to create heavy and light chain immunoglobulin cDNA expression libraries in the λ ImmunoZap(H) and λ ImmunoZap(L) vectors. These vectors may be screened individually or co-expressed to form Fab fragments or antibodies (*see* Huse et al.,
5 *supra*; *see also* Sastry et al., *supra*). Positive plaques may subsequently be converted to a non-lytic plasmid which allows high level expression of monoclonal antibody fragments from *E. coli*.

Similarly, portions or fragments, such as Fab and Fv fragments, of antibodies may also be constructed utilizing conventional enzymatic digestion or
10 recombinant DNA techniques to incorporate the variable regions of a gene which encodes a specifically binding antibody. Within one embodiment, the genes which encode the variable region from a hybridoma producing a monoclonal antibody of interest are amplified using nucleotide primers for the variable region. These primers may be synthesized by one of ordinary skill in the art, or may be purchased from
15 commercially available sources. Stratacyte (La Jolla, Calif.) sells primers for mouse and human variable regions including, among others, primers for $V_{H\alpha}$, $V_{H\beta}$, $V_{H\gamma}$, $V_{H\delta}$, C_{H1} , V_L and C_L regions. These primers may be utilized to amplify heavy or light chain variable regions, which may then be inserted into vectors such as ImmunoZAP™ H or ImmunoZAP™ L (Stratacyte), respectively. These vectors may then be introduced into
20 *E. coli*, yeast, or mammalian-based systems for expression. Utilizing these techniques, large amounts of a single-chain protein containing a fusion of the V_H and V_L domains may be produced (*see* Bird et al., *Science* 242:423-426, 1988). In addition, such techniques may be utilized to change a "murine" antibody to a "human" antibody, without altering the binding specificity of the antibody.

25 Once suitable antibodies have been obtained, they may be isolated or purified by many techniques well known to those of ordinary skill in the art (*see Antibodies: A Laboratory Manual*, Harlow and Lane (eds), Cold Spring Harbor

30 combination of these techniques.

Assays

Assays useful within the context of the present invention include those assays for detecting agonists or antagonists of WRN protein activity. Other assays are useful for the screening of peptide or organic molecule libraries. Still other assays are useful for the identification and/or isolation of nucleic acid molecules and/or peptides within the present invention, the identification of proteins that interact or bind the WRN protein, for diagnosis of a patient with an increased likelihood of contracting Werner's Syndrome, or for diagnosis of a patient with susceptibility to or manifestation of a WRN-related disease.

Nucleic Acid Based Diagnostic Tests

Briefly, another aspect of the present invention provides probes and primers for detecting the WRN genes and/or mutants thereof. In one embodiment of this aspect, probes are provided that are capable of specifically hybridizing to DNA or RNA of the WRN genes. For purposes of the present invention, probes are "capable of hybridizing" to DNA or RNA of the WRN gene if they hybridize to an WRN gene under conditions of either high or moderate stringency (*see* Sambrook et al., *supra*) but not significantly or detectably to the an unrelated helicase gene such as the Bloom's Syndrome gene (Ellis et al., *Cell* 83:655-666, 1995). Preferably, the probe hybridizes to suitable nucleotide sequences under high stringency conditions, such as hybridization in 5x SSPE, 1x Denhardt's solution, 0.1% SDS at 65°C, and at least one wash to remove unhybridized probe in the presence of 0.2x SSC, 1x Denhardt's solution, 0.1% SDS at 65°C. Except as otherwise provided herein, probe sequences are designed to allow hybridization to WRN genes, but not to DNA or RNA sequences from other genes. The probes are used, for example, to hybridize to nucleic acid that is present in a biological sample isolated from a patient. The hybridized probe is then detected, thereby detecting the presence of the WRN gene in the sample. Alternatively, the WRN gene may be amplified and the amplified

product subjected to DNA sequencing. Mutants of WRN may be detected by DNA sequence analysis or hybridization with allele-specific oligonucleotide probes under conditions and for time sufficient to allow hybridization to the specific allele. Typically, the hybridization buffer and wash will contain tetramethyl ammonium chloride or the like (*see* Sambrook et al., *supra*).

Nucleic acid probes of the present invention may be composed of either deoxyribonucleic acids (DNA), ribonucleic acids (RNA), nucleic acid analogues (*e.g.*, peptide nucleic acids), or any combination thereof, and may be as few as about 12 nucleotides in length, usually about 14 to 18 nucleotides in length, and possibly as large as the entire sequence of a WRN gene. Selection of probe size is somewhat dependent upon the use of the probe, and is within the skill of the art.

Suitable probes can be constructed and labeled using techniques that are well known in the art. Shorter probes of, for example, 12 bases can be generated synthetically and labeled with ^{32}P using T₄ polynucleotide kinase. Longer probes of about 75 bases to less than 1.5 kb are preferably generated by, for example, PCR amplification in the presence of labeled precursors such as [α - ^{32}P]dCTP, digoxigenin-dUTP, or biotin-dATP. Probes of more than 1.5 kb are generally most easily amplified by transfecting a cell with a plasmid containing the relevant probe, growing the transfected cell into large quantities, and purifying the relevant sequence from the transfected cells. (*See* Sambrook et al., *supra*.)

Probes can be labeled by a variety of markers, including for example, radioactive markers, fluorescent markers, enzymatic markers, and chromogenic markers. The use of ^{32}P is particularly preferred for marking or labeling a particular probe.

It is a feature of this aspect of the invention that the probes can be utilized to detect the presence of WRN mRNA or DNA within a sample. However, if the relevant sample is present in only a limited number, then it may be beneficial to

sequence including, for example, RNA amplification (*see* Lizardi et al.,

Bio/Technology 6:1197-1202, 1988; Kramer et al., *Nature* 339:401-402, 1989; Lomeli et al., *Clinical Chem.* 35(9):1826-1831, 1989; U.S. Patent No. 4,786,600), and DNA amplification utilizing LCR or polymerase chain reaction ("PCR") (see, U.S. Patent Nos. 4,683,195, 4,683,202, and 4,800,159) (see also U.S. Patent Nos. 4,876,187 and 5,011,769, which describe an alternative detection/amplification system comprising the use of scissile linkages), or other nucleic acid amplification procedures that are well within the level of ordinary skill in the art. With respect to PCR, for example, the method may be modified as known in the art. Transcriptional enhancement of PCR may be accomplished by incorporation of bacteriophage T7 RNA polymerase promoter sequences in one of the primary oligonucleotides, and immunoenzymatic detection of the products from the enhanced emitter may be effected using anti-RNA:DNA antibodies (Blais, *Appl. Environ. Microbiol.* 60:348-352, 1994). PCR may also be used in combination with reverse dot-blot hybridization (Iida et al., *FEMS Microbiol. Lett.* 114:167-172, 1993). PCR products may be quantitatively analyzed by incorporation of dUTP (Dupl  a et al., *Anal. Biochem.* 212:229-236, 1993), and samples may be filter sampled for PCR-gene probe detection (Bej et al., *Appl. Environ. Microbiol.* 57:3529-3534, 1991).

Within a particularly preferred embodiment, PCR amplification is utilized to detect the WRN DNA. Briefly, as described in greater detail below, a DNA sample is denatured at 95  C in order to generate single-stranded DNA. The DNA sample may be a cDNA generated from RNA. Specific primers are then annealed to the single-stranded DNA at 37  C to 70  C, depending on the proportion of AT/GC in the primers. The primers are extended at 72  C with *Taq* DNA polymerase or other thermostable DNA polymerase in order to generate the opposite strand to the template. These steps constitute one cycle, which may be repeated in order to amplify the selected sequence. For greater specificity, nested PCR may be performed. In nested PCR, a second amplification is performed using a second set of primers derived from sequences

convenient size for determining their sequence. In a preferred embodiment, nested PCR is performed.

Within an alternative preferred embodiment, LCR amplification is utilized for amplification. LCR primers are synthesized such that the 5' base of the upstream primer is capable of hybridizing to a unique base pair in a desired gene to specifically detect an WRN gene.

Within another preferred embodiment, the probes are used in an automated, non-isotopic strategy wherein target nucleic acid sequences are amplified by PCR, and then desired products are determined by a colorimetric oligonucleotide ligation assay (OLA) (Nickerson et al., *Proc. Natl. Acad. Sci. USA* 81:8923-8927, 1990).

Primers for the amplification of a selected sequence should be selected from sequences that are highly specific to WRN (and not, e.g., the Bloom's Syndrome gene, *supra*) and form stable duplexes with the target sequence. The primers should also be non-complementary, especially at the 3' end, should not form dimers with themselves or other primers, and should not form secondary structures or duplexes with other regions of DNA. In general, primers of about 18 to 20 nucleotides are preferred, and can be easily synthesized using techniques well known in the art. PCR products, and other nucleic acid amplification products, may be quantitated using techniques known in the art (Dupl   et al., *Anal. Biochem.* 212:229-236, 1993; Higuchi et al., *Bio/Technology* 11:1026-1030).

Within one embodiment of the invention, nucleic acid diagnostics may be developed which are capable of detecting the presence of Werner's Syndrome, or of various related diseases that may be caused by Werner's Syndrome. Briefly, severe mutations in the WRN gene may lead to Werner's Syndrome, as well as a host of related diseases, including for example, increased frequency of some benign and malignant neoplasms (especially sarcomas), cataracts, cardiovascular disease, osteoporosis, type I

diabetes. In addition, many of the related diseases may be associated with mutations in

the WRN gene. For example, diabetes and osteoporosis are often associated with aging. Aging population and individuals with these (or other) diseases are screened for mutations in WRN. Any of the assays described herein may be used. RT-PCR is especially preferred in conjunction with DNA sequence determination. To correlate a mutation or polymorphism with disease, sibling pairs in which one sibling has disease are preferred subjects. Once a mutation is identified, other convenient screening assays may be used to assay particular nucleotide changes.

Since the sequences of the two copies of the gene from non-Werner's affected individuals can be correlated with the medical histories of these patients to define these correspondences, these alleles can therefore be used as diagnostics for susceptibilities to these diseases, once the relationship is defined. Certain non-null forms of the gene, for example, in either the homozygous or heterozygous state may significantly affect the propensity for the carriers to develop, for example, cancer. These propensities can be ascertained by examining the sequences of the gene (both copies) in a statistically significant sample of cancer patients. Other diseases (see above) can be similarly examined for significant correlations with certain alleles. To detect such a causal relationship one can use a chi-squared test, or other statistical test, to examine the significance of any correlation between the appropriate genotypes and the disease state as recorded in the medical records, using standard good practices of medical epidemiology. The sequences that define each of the alleles are then valuable diagnostic indicators for an increased susceptibility to the disease. Thus, from the nucleic acid sequences provided herein, a wide variety of Werner's Syndrome-related diseases may be readily detected.

Another cellular phenotype of the cells from Werner's patients is the increased frequency of deletion mutation in these cells. Clearly, the defective helicase in these cells leads to a specific mutator phenotype, while not rendering the cells hypersensitive to a variety of chemical or physical mutagens that damage DNA, like

some alleles may therefore be diagnostic of this class of medical conditions.

Assays for agonists and antagonists

Also provided by the present invention are agonists or antagonists of the WRN gene product comprising a protein, peptide, chemical, or peptidomimetic that
5 binds to the WRN gene product or interacts with a protein that binds to the WRN gene product such that the binding of the agonist or antagonist affects the activity of the WRN gene product. An agonist will activate or increase the activity of the WRN gene product. An antagonist will inhibit or decrease the activity of the WRN gene product. The activity of the WRN gene product may be measured in an assay, such as a helicase
10 assay or other assay that measures an activity of the WRN gene product. Other assays measure the binding of protein that interacts with WRN and is necessary for its activity.

Agonists and antagonists of the WRN gene product may be used to enhance activity or inhibit activity of the gene product. Such agonists and antagonists may be identified by a variety of methods. For example, proteins that bind and activate
15 WRN may be identified using a yeast 2 hybrid detection system. In this system, the WRN gene is fused to either a DNA-binding domain or an activating domain of a yeast gene such as GAL4. A cDNA library is constructed in a vector such that the inserts are fused to one of the domains. The vectors are co-transfected into yeast and selected for transcriptional activation of a reporter gene (Fields and Song, *Nature* 340: 245, 1989).
20 The protein(s) that bind to WRN are candidate agonists. Three different proteins that bind WRN have been identified in an initial screen using the 2-hybrid system.

When the binding site on WRN gene product is determined, molecules that bind and activate WRN protein can be designed and evaluated. For example, computer modeling of the binding site can be generated and mimetics that bind can be
25 designed. Antibodies to the binding site may be generated and analogues of native binding proteins generated as well. Any of these molecules is tested for agonist or antagonist activity by a functional assay of the WRN gene product. For example, to test

administered and activation is monitored. An antagonist will inhibit the activation of

the reporter gene by at least 50%. Similarly, agonist activity may be measured by either enhancing WRN activity in a yeast 2-hybrid system or by coupling the test compound to a DNA binding or activation domain and monitoring activity of the reporter gene.

Labels

WRN proteins, nucleic acid molecules which encodes such proteins, anti-WRN protein antibodies and agonists or antagonists, as described above and below, may be labeled with a variety of molecules, including for example, fluorescent molecules, toxins, and radionuclides. Representative examples of fluorescent molecules include fluorescein, *Phycobili* proteins, such as phycoerythrin, rhodamine, Texas red and luciferase. Representative examples of toxins include ricin, abrin diphtheria toxin, cholera toxin, gelonin, pokeweed antiviral protein, tritin, *Shigella* toxin, and *Pseudomonas* exotoxin A. Representative examples of radionuclides include Cu-64, Ga-67, Ga-68, Zr-89, Ru-97, Tc-99m, Rh-105, Pd-109, In-111, I-123, I-125, I-131, Re-186, Re-188, Au-198, Au-199, Pb-203, At-211, Pb-212 and Bi-212. In addition, the antibodies described above may also be labeled or conjugated to one partner of a ligand binding pair. Representative examples include avidin-biotin, and riboflavin-riboflavin binding protein.

Methods for conjugating or labeling the WRN proteins, nucleic acid molecules which encode such proteins, anti-WRN protein antibodies and agonists or antagonists, as discussed above, with the representative labels set forth above may be readily accomplished by one of ordinary skill in the art (see Trichothecene Antibody Conjugate, U.S. Patent No. 4,744,981; Antibody Conjugate, U.S. Patent No. 5,106,951; Fluorogenic Materials and Labeling Techniques, U.S. Patent No. 4,018,884; Metal Radionuclide Labeled Proteins for Diagnosis and Therapy, U.S. Patent No. 4,897,255; and Metal Radionuclide Chelating Compounds for Improved Chelation Kinetics, U.S. Patent No. 4,988,496; see also Inman, *Methods In Enzymology*, Vol. 34, "Affinity

Bioanalytical Applications," *Anal. Biochem.* 171:1-32, 1988).

Pharmaceutical Compositions

As noted above, the present invention also provides a variety of pharmaceutical compositions, comprising one of the above-described WRN proteins, nucleic acid molecules, vectors, antibodies, host cells, agonists or antagonists, along with a pharmaceutically or physiologically acceptable carrier, excipients or diluents. Generally, such carriers should be nontoxic to recipients at the dosages and concentrations employed. Ordinarily, the preparation of such compositions entails combining the therapeutic agent with buffers, antioxidants such as ascorbic acid, low molecular weight (less than about 10 residues) polypeptides, proteins, amino acids, carbohydrates including glucose, sucrose or dextrans, chelating agents such as EDTA, glutathione and other stabilizers and excipients. Neutral buffered saline or saline mixed with nonspecific serum albumin are exemplary appropriate diluents.

In addition, the pharmaceutical compositions of the present invention may be prepared for administration by a variety of different routes. In addition, pharmaceutical compositions of the present invention may be placed within containers, along with packaging material which provides instructions regarding the use of such pharmaceutical compositions. Generally, such instructions will include a tangible expression describing the reagent concentration, as well as within certain embodiments, relative amounts of excipient ingredients or diluents (*e.g.*, water, saline or PBS) which may be necessary to reconstitute the pharmaceutical composition.

Methods of Treating or Preventing Werner's Syndrome

The present invention also provides methods for treating or preventing Werner's Syndrome (or related diseases), comprising the step of administering to a patient a vector (*e.g.*, expression vector, viral vector, or viral particle containing a vector) or nucleic acid molecules alone, as described above, thereby reducing the risk of developing Werner's Syndrome. The present invention also provides a method for using the present invention to delay onset of Werner's Syndrome, lessen symptoms, or hasten or delay

progression of the disease. Such therapeutics may be tested in a transgenic animal model that expresses mutant protein, wild-type and mutant protein, or in an *in vitro* assay system (e.g., a helicase assay such as that described by Bjornson et al., *Biochem. 3307:14306-14316*, 1994).

5 As noted above, the present invention provides methods for treating or preventing Werner's Syndrome through the administration to a patient of a therapeutically effective amount of an antagonist or pharmaceutical composition as described herein. Such patients may be identified through clinical diagnosis based on the classical symptoms of Werner's Syndrome.

10 As will be evident to one of skill in the art, the amount and frequency of administration will depend, of course, on such factors as the nature and severity of the indication being treated, the desired response, the condition of the patient, and so forth. Typically, the compositions may be administered by a variety of techniques, as noted above.

15 Within other embodiments of the invention, the vectors which contain or express the nucleic acid molecules which encode the WRN proteins described above, or even the nucleic acid molecules themselves may be administered by a variety of alternative techniques, including for example administration of asialoosomucoid (ASOR) conjugated with poly-L-lysine DNA complexes (Cristano et al., *PNAS* 92:122-
20 92126, 1993), DNA linked to killed adenovirus (Curiel et al., *Hum. Gene Ther.* 3(2):147-154, 1992), cytofectin-mediated introduction (DMRIE-DOPE, Vical, Calif.), direct DNA injection (Acsadi et al., *Nature* 352:815-818, 1991); DNA ligand (Wu et al., *J. of Biol. Chem.* 264:16985-16987, 1989); lipofection (Felgner et al., *Proc. Natl. Acad. Sci. USA* 84:7413-7417, 1989); liposomes (Pickering et al., *Circ.* 89(1):13-21, 1994;
25 and Wang et al., *PNAS* 84:7851-7855, 1987); microprojectile bombardment (Williams et al., *PNAS* 88:2726-2730, 1991); and direct delivery of nucleic acids which encode the WRN protein itself either alone (Vile and Hart, *Cancer Res.* 53: 3860-3864, 1993) or

The following examples are offered by way of illustration, and not by way of limitation.

EXAMPLES

EXAMPLE 1

CLONING OF THE WRN GENE FROM CHROMOSOME 8

5

The WS locus (*WRN*) was initially localized to 8p12 by conventional mapping methods (Goto et al., *Nature* 355:735-738, 1992) and the genetic position refined using both meiotic and homozygosity mapping (Schellenberg et al., 1992; Nakura, et al., *Genomics* 23:600-608, 1994; Thomas, *Genomics* 16:685-690, 1993).

10 The latter approach is possible since many WS subjects are the offspring of consanguineous marriages (Table 1). Initial mapping work (Nakura, et al., *Genomics* 23:600-608, 1994; Oshima et al., *Genomics* 23:100-113, 1994) placed the *WRN* locus in an 8.3 cM interval flanked by *D8S137* and *D8S87* (Fig. 1). *D8S339*, a marker within this interval, was the closest locus tested ($q = 0.001$, $Z_{\max} = 15.93$). Multipoint analysis
15 placed *WRN* within 0.6 cM of *D8S339*, although the region between *D8S87* and *FGFR* could not be excluded. Subsequently, the short tandem repeat polymorphism (STRP) markers at glutathione reductase (*GSR*) and *D8S339* were found to be in linkage disequilibrium with WS in Japanese WS subjects (Yu, *American Journal of Human Genetics* 55:356-364, 1994).

20 To clone the *WRN* gene, a yeast artificial chromosome (YAC) P1, and cosmid contig was generated starting at the *GSR/D8S339* region and extended by walking methods to cover approximately 3 Mb. An additional 16 STRP markers in the YAC contig (Fig. 1B) were identified to define recombinants and to delineate the boundaries of the linkage disequilibrium region. For marker ordering and gene
25 identification, cosmids and P1 clones were also isolated and used to construct a small-clone partial contig of the region (Fig. 1E). The *WRN* region was defined by obligate recombinants at C41C3S3 excluding the region telomeric to this marker, and at C86P9

30 region

Genes in the WRN region were identified by exon trapping using vector pSL3 (Buckler et al., *Proc. Natl. Acad. Sci. USA* 88:4005-4009, 1991; Church et al., *Nat. Genet.* 6:98-105, 1994), hybridization of cDNA libraries to immobilized YACs (Parimoo et al., *Proc. Natl. Acad. Sci. USA* 87:3166-3169, 1991), and comparison of the genomic sequence to DNA sequence databases using BLAST (Altschul et al. *J. Mol. Biol.* 215:403-410, 1990) and the exon-finding program GRAIL (Uberbacher and Mural, *Proc. Natl. Acad. Sci. USA* 88:1261, 1991). The genomic sequence was determined for the region defined by P1 clones 2233, 2253, 3833, 2236, 2237, 2932, 6738 and 2934 and cosmid clone 176 C6. Each method identifies short segments of expressed sequences, which were then used to screen an arrayed fibroblast cDNA library to identify longer cDNA clones. This library was selected because WS fibroblasts have a premature senescence phenotype *in vitro*, indicating that the WRN gene is probably expressed in this cell type. Genes identified by this process were screened for WRN mutations using reverse transcriptase-polymerase chain reaction (RT-PCR). Seven subjects were initially screened for mutations: 5 WRN subjects (2 Caucasians and 3 Japanese) and 2 control subjects (1 Caucasian and 1 Japanese). Prior to identification of the WRN gene, the following genes from the region were screened for mutations: GSR, PP2AB, TFIIEB, and genes corresponding to other expressed sequence tagged sites (ESTs).

The candidate WRN locus gene was initially detected by using the genomic sequence of P1 clone 2934 to search the EST database. A single 245 bp EST, R58879, was detected which is homologous to 3 segments of the genomic sequence separated by presumed intronic sequence. Sequence from R58879 was used to identify longer cDNA clones from a normal fibroblast cDNA library. An initial 2.1 kb cDNA clone containing EST R58879, which corresponds to the 3' end of the gene, was obtained by screening an array of clones by PCR, using the primers A and B (see below). Primers A and B are derived from R58879 sequence and yield a 145 bp

located in p2934 and to sequences contained in the initial 2.1 kb clone. Six additional

clones were identified. An additional 8 clones were obtained by plaque hybridization. The longest clone is 4.0 kb in length. Additional sequence was obtained by the RAGE method using primer 5EA to prime first strand cDNA synthesis. A 2.5 kb product was obtained that contained an additional 1.4 kb of sequence.

5 Evidence that R58879 is expressed was obtained by Northern blot analysis, in which 6.5 kb and 8 kb transcripts were detected in a variety of tissues, including heart, placenta, muscle, and pancreas. Also, transcripts were detected by RT-PCR products from fibroblast and lymphoblastoid cell line RNA.

10 EXAMPLE 2

CLONING OF THE WRN GENE FROM SUBJECTS

The WRN gene may be isolated from patients and mutations or polymorphisms determined by sequence analysis. Peripheral blood cells are obtained
15 by venipuncture and hypotonic lysis of erythrocytes. DNA or RNA is isolated from these cells and the WRN gene isolated by amplification. The gene sequence may be obtained by amplification of the exons from genomic DNA or by RT-PCR, followed by determination of the DNA sequence. Primers suitable for determining the DNA sequence and for performing RT-PCR are listed below (Primers A-R are SEQ ID Nos.
20 1-18 respectively, and primers 5EA-5EG are SEQ ID Nos. 19-25 respectively). Two cDNAs were identified and are shown in Figures 2 and 3. There is some uncertainty regarding the identity of a few bases in the 5' untranslated region in Figure 2

Two RT-PCR reactions are used to obtain the gene from different tissues. First strand cDNA synthesis is carried out according to standard procedures
25 (e.g., with a Stratascript Kit from Stratagene). The cDNA is subjected to a pair of nested PCR amplifications, the first with primers I and J (SEQ ID Nos. 9 and 10), followed by primers K and L (SEQ ID Nos. 11 and 12), and the second with primers

30

the gene sequence or splicing pattern. Primers A-H (SEQ ID Nos. 1-8) and K-R (SEQ

ID Nos. 11-18) are used for sequencing the first RT-PCR fragment. Primers B, 5EA, 5EB, 5EC, 5EE, 5EF and 5EG (SEQ ID Nos. 2, 19, 20, 21, 23, 24, and 25, respectively) are used for sequencing the second RT-PCR fragment. Sequencing is done on an ABI373A using Applied Biosystems Division of Perkin-Elmer FS sequencing kits according to the instructions of the manufacturer.

A	5'-CTGGCAAGGATCAAACAGAGAG
B	5'-CTTTATGAAGCCAATTTCTACCC
C	5'-TGGCAAATTGGTAGAAGCTAGG
10 D	5'-AAATAACTATGCTTTCTTACATTTAC
F	5'-CTCCCGTCAACTCAGATATGAG
F	5'-CTGTTTGTAATGTAAGAAAGCATAG
G	5'-GAGCTATGATGACACCACTGC
H	5'-ACTGAGCAACAGAGTGAGACC
15 I	5'-GGATCTGGTCTCACTCTGTTGC
J	5'-TTGCCTAGTGCAATTGGTCTCC
K	5'-AGTGCAGTGGTGTATCATAGC
L	5'-CCTATTTAATGGCACCCAAAATGC
M	5'-CAGTCTATGGCCATCACATACTC
20 N	5'-ACCGCTTGGGATAAGTGCATGC
O	5'-GAGAAGAAGTCTAACTTGGAGAAG
P	5'-TTCTGGTGACTGTACCATGATAC
Q	5'-CCAAAGGAAGTGATACCAGCAAG
R	5'-ACAGCAAGAAACATAATTGTTCTGG
25 5EA	5'-GAACTTTGAAGTCCATCACGACC
5EB	5'-GCATTAATAAAGCTGACATTCGCC
5EC	5'-CATTACGGTGCTCCTAAGGACATG
5ED	5'-GATGGATTTGAAGATGGAGTAGAAG
5EE	5'-TGAAAGACAATATGGAAAGAGCTTG
30 5EF	5'-GTAGAACCAACTCATTCTAAATGCT
5EG	5'-AATTGCGTGTCTATCCTTGCGCA

The exons of the 3'-end of the WRN gene can be amplified from DNA samples using the primers listed below (Primers E1A-E13B are SEQ ID Nos. 26-57, respectively). The DNA sequence is determined using the same primers and an

E1A 5'-CTTAACTCAAGGATGGAAGT

```

E1B 5'-CATGAAACTTGCTTCTAGGACAC
E2A 5'-CCCAGGAGTTGAGACCATCC
E2B 5'-TTACAATCGGCCACATTCATCAC
E2C 5'-TGTAATCCCAACACTTTGGGAGG
5  E2D 5'-AGTGAAGAATTCATAGTGGATGG
E3A 5'-TAGCTTTATGAAGCCAATTTCTACC
E3B 5'-AATCCAAAGAATCAATAGACAAGTC
E3C 5'-GCTTGAAGGATGAGGCTCTGAG
E3D 5'-TGTCAGAAATGAGCACCATGGG
10 E4A 5'-CTTGTGAGAGGCCTAATAAAGTCG
E4B 5'-GGTAAACAGTGTAGGAGTCTGC
E5A 5'-GCCATTTTCTCTTAATTGGAAAGG
E5B 5'-ATCTTATTCATCTTTCTGAGAATGG
E6A 5'-TGAAATAGCCCAACATCTGACAG
15 E6B 5'-GATTAATTTGACAGCTTGATTAGGC
E7A 5'-TGAAATATAAACTCAGACTCTTAGC
E7B 5'-GTACTGATTTGGAAAGACATTCTC
E8A 5'-GATGTGACAGTGAAGCTATGG
E8B 5'-GGAAAAATGIGGTATCTGAAGCTC
20 E9A 5'-AAGTGAGCAATGTTGCTTCTGG
E9B 5'-TCATTAGGAAGCTGAACATCAGC
E10A 5'-GTTGGAGGAAATTGATCCCMAGTC
E10B 5'-TGTTGCTTATGGGTTTAAGTTGTG
E11A 5'-TAAAGGATTAATGCTGTTAACAGTG
25 E11B 5'-TCACACTGAGCATTACTACCTG
E12A 5'-GTAAATCATAATCAGAATTCATAACAG
E12B 5'-CTTTGGCAACCTTCCACCTTCC
E12C 5'-GCAAAAGGAAATGTAGCACATAGAG
E12D 5'-AGGCTATAGGCATTTGAAAGAGG
30 E13A 5'-GTAGGCTCCAGAAAGACCCAG
E13B 5'-GAAAGGATGGGTGTGTATTCAGG

```

EXAMPLE 3

35

IDENTIFICATION OF MUTANT ALLELES

The cDNA sequence (Figure 2) was aligned to the genomic sequence to

identify the mutations.

In the first four patients, single base pair changes lead to splicing defects or stop codons in the open reading frame of the gene. In the fifth patient, a single base pair

change results in a cysteine to arginine transition, which may disrupt gene function. Each of the exons was also sequenced in 96 unaffected control individuals (48 Caucasians and 48 Japanese), and none of the mutations were found in any of the control individuals.

- 5 The first mutation is a mutation at a splice acceptor site. In the sequence below, the GGTAGAAA sequence begins at nucleotide 2030 (Figure 2). The g to c change results in a deletion of 95 bp.

- Preparation of DNA for RT-PCR mutational analysis revealed that for one subject, the amplification product was shorter than observed in products from other
10 WS and control subjects. DNA sequence analysis of the RT-PCR product revealed that 95 bp were missing compared to other samples. The missing sequence corresponds to a single exon. This exon and flanking genomic segments were sequenced from the WS subject and controls and a single base change (G→C) at the splice donor site was detected. The subject was the offspring of a first cousin marriage and was, as expected,
15 homozygous for this mutation. The same mutation was found in a total of 18 out of 30 Japanese WS subjects and, thus, is the most common Japanese WS mutation. Deletion of this exon results in a change in the predicted open-reading frame and a premature stop codon. This mutation was not observed in 46 Japanese and 46 Caucasian controls. Among mutation carriers, 12/16 had the 141 bp allele at the GSR2-STRP.

20

wild type: ttttaatatagGGTAGAAA	(SEQ ID No. 58)
Werners: ttttaatacGGTAGAAA	(SEQ ID No. 59)

- The second mutation changes a C to T at nucleotide 2384 (Figure 2)
25 changing a glutamine to a stop codon, which results in a predicted truncated protein. This mutation was observed in a single subject. Primers E11A and E11B flank this sequence and amplify a 360 bp fragment.

primer	5'	3'	stop
		<u>T</u>	
		ter	

The third mutation changes a C to T at nucleotide 2804 (Figure 2), which alters an arginine codon to a stop codon resulting in a predicted truncated protein. Four Japanese WS subjects and 1 Caucasian W5 subject had this mutation. Primers E8A and E8B flank this sequence and amplify a 267 bp product.

arg
wild type: TTGGAGCGAGCA (SEQ ID No. 62)
Werners: TTGGAGTGAGCA (SEQ ID No. 63)
ter

The fourth mutation is a 4 bp deletion across a splice junction. The exon sequence shown below begins at nucleotide 2579 (Figure 2). This mutation was identified in a Syrian W5 kindred. Primers E4A and E4B flank this mutation and amplify a 267 bp fragment.

wild type: ctgt**ag**ACAGACACCTC (SEQ ID No. 68)
Werners: ctgt----AGACACCTC (SEQ ID No. 69)

The fifth mutation is a missense mutation. A T is altered to a G at nucleotide 2113 (Figure 2), changing the wild-type phe codon to a leu codon. This change is a polymorphism with each allele present at a frequency of approximately 0.5. It does not appear to correlate with WS.

phe
wild type: AAGAAGTTTCTTCTG (SEQ ID No. 64)
Werners: AAGAAGTTGCTTCTG (SEQ ID No. 65)
leu

The sixth mutation is a missense mutation changing a T to a C at

wild type: CTTCATGTCAT (SEQ ID No. 66)
Werners: CTTCACTGAT (SEQ ID No. 67)

arg

These point mutations may also be identified by PCR using primers that contain as the 3'-most base either the wild type or the mutant nucleotide. Two separate
5 reactions are performed using one of these primers and a common second primer. Amplification is detectable in the reaction containing a matched primer.

EXAMPLE 4

10 CHARACTERIZATION OF THE WRN GENE AND GENE PRODUCT

The 2 kb WRN cDNA hybridizes to a 6.5 kb RNA and a less abundant 8 kb RNA on a Northern blot, suggesting that a full length coding region is about 5.2 kb long. An overlapping cDNA clone has been isolated that extends the sequence by 2 kb.
15 The insert from this clone is used to probe cDNA libraries to identify other clones that contain the 5' end of the cDNA or full length sequence. Alternate splicing events are detected by sequencing the full cDNA sequence from a number of different tissues, including fully differentiated cells and stem cells, and the full range of gene transcripts identified by sequence comparison. Additional exons are identified as above by further
20 genomic sequencing and GRAIL analysis.

The predicted amino acid sequence is shown in Figures 2B and 3. Figure 2 shows cDNA and predicted amino acid sequences of the WRN gene. Figure 3 presents cDNA and predicted amino acid sequences of a less abundant transcript of the WRN gene. The longest open reading frame is shown from the first methionine in that
25 frame. The predicted WRN protein consists of 1,432 amino acids divided into three regions: an N-terminal region, a central region containing 7 motifs (I, Ia, II, III, IV, V and VI) characteristic of the DNA and RNA superfamily of helicases (Gorbalenya et al. *Nucleic Acid Res.* 17: 1713-1989, 1989; Gorbalenya et al. *EMBO J.* 8: 1473-1481, 1989).

Now, having described the WRN gene and the WRN protein, we now describe the WRN protein. Because many helicases function as part of a multiprotein complex, the N-terminal

and/or the C-terminal domain may contain interaction sites for these other proteins, while the central helicase domain functions in the actual enzymatic unwinding of DNA or RNA duplexes.

The N-terminal region, encompassing approximately codons 1 to 539, is
5 acidic; there are 109 aspartate or glutamate residues, including a stretch of 14 acidic residues in a 19 amino acid sequence (codons 507-526). Stretches of acidic residues are found in the Xeroderma pigmentosum (XP) complementation group B helicase, the Bloom's syndrome helicase, and the X-chromosome-linked α -thalassemia mental retardation syndrome helicase. In the WRN gene, this region also contains a tandem
10 duplication of 27 amino acids in which each copy is encoded by a single exon. Because this duplication is exact at the nucleotide level, and because flanking intronic sequences for the two exons that encode the duplication are also highly similar, this duplication is presumed to be the result of a relatively recent event. The duplicated regions are also highly acidic with 8 glutamate or aspartate residues out of 27 amino acids and only 2
15 basic amino acids (one histidine and one lysine residue).

The central region of the WRN gene, spanning approximately codons 540-963, is highly homologous to other helicases from a wide range of organisms including the ReqQ gene from *E. coli*, the SGS1 gene from *S. cerevisiae*, a predicted helicase (F18C5C) from *C. elegans*, and several human helicases. Thus, by sequence
20 similarity, the WRN gene is a member of a superfamily of DExH-box DNA and RNA helicases. The principle conserved sequences consist of 7 motifs found in other helicases. These motifs include a predicted nucleotide binding site (motif I) and a Mg²⁺ binding site (sequence DEAH, motif II). Some or all of the 7 motifs are presumed to form the enzymatic active site for DNA/RNA unwinding. The presence of the DEAH
25 sequence and an ATP-binding motif further suggests that the WRN gene product is a functional helicase.

The C-terminal end of the WRN gene, from codons 964 to 1432, has

EXAMPLE 5

IDENTIFYING AND DETECTING MUTATIONS IN THE WRN GENE

5

Mutations or polymorphisms of WRN may be identified by various methods, including sequence analysis. Although any cell (other than erythrocytes) may be used to isolate nucleic acids, peripheral blood mononuclear cells (PBMC) are preferred. Peripheral blood mononuclear cells are obtained by venipuncture and subsequent hypotonic lysis of erythrocytes. RNA is isolated and first strand cDNA synthesis is performed using a Strata-script RT-PCR kit according to the manufacturers instructions (Stratagene, La Jolla, part numbers 200347 and 200420). Three RT-PCR fragments are amplified using an LA PCR Kit Ver. 2 using buffer containing 1.5 mM Mg+2 (TaKaRa Shuzo Co., Ltd., Japan, part number RR013A). Nested PCR is performed. In this reaction, a second PCR is performed using a pair of primers within the sequence amplified by the first PCR reaction. The cycling conditions for each amplification are: 10 min at 95°C, 35 cycles of 1 min at 60°C, 1 min at 72°C, and 1 min at 95°C, followed by 7 min at 72°C in a Perkin-Elmer 9600 PCR machine. The amplified fragments are purified using 96-well plate spin columns (Wang et al., *Anal. Biochem.* 226:85-90, 1995). DNA sequence is determined using an FS Dye-Terminator sequencing kit (Applied Biosystems Division of Perkin Elmer) and the specific primers described below. An automated Applied Biosystems ABI373A DNA Sequencer is used to determine the sequence. The amplified fragments and the appropriate primers are listed in Table 1, and the primer sequences are listed in Table 2.

25

The DNA sequences are aligned with the known sequence (Figure 2A) using the program Sequencher (Gene Codes, Michigan) to identify any discrepancies between patient samples and the reference sequence.

Table 1 PCR and sequence primers

Fragment	Primers Nested on cDNA		Coordinates	Sequence primers
	1st PCR	2nd PCR		
I	5EC, J	5EN, L	2947-5065	5EN, L, M, N, O, P, Q, R
II	5ED, P	5EE, B	1379-3391	5EE, 5EJ, 5EK, 5EL, 5EM, 5EB, 5EA, 5EN, B
III	5ES, 5EK	5ET, 5EH	75-1516	5ET, 5EX, 5E1, 5EP, 5EO, 5ED, 5EH

Table 2 Primer sequences

5	B	5'-CTTTATGAAGCCAATTTCTACCC	(SEQ ID No. 2)
	J	5'-TTGCCTAGTGCAATTGGTCTCC	(SEQ ID No. 10)
	L	5'-CCTATTTAATGGCACCACAAATGC	(SEQ ID No. 12)
	M	5'-CAGTCTATGGCCATCACATACTC	(SEQ ID No. 13)
10	N	5'-ACCGCTTGGGATAAGTGCATGC	(SEQ ID No. 14)
	O	5'-GAGAAGAAGTCTAACTTGGAGAAG	(SEQ ID No. 15)
	P	5'-TTCTGGTGACTGTACCATGATAC	(SEQ ID No. 16)
	Q	5'-CCAAAGGAAGTGATACCAGCAAG	(SEQ ID No. 17)
	R	5'-ACAGCAAGAAACATAATTGTTCTGG	(SEQ ID No. 18)
15	5EA	5'-GAACTTTGAAGTCCATCACGACC	(SEQ ID No. 19)
	5EB	5'-GCATTAATAAAGCTGACATTCGCC	(SEQ ID No. 20)
	5EC	5'-CATTACGGTGCTCCTAAGGACATG	(SEQ ID No. 21)
	5ED	5'-GATGGATTTGAAGATGGAGTAGAAG	(SEQ ID No. 22)
	5EE	5'-TGAAAGAGAATATGGAAAGAGCTTG	(SEQ ID No. 23)
20	5EH	5'-CATTGGGAGATAAATGCTCAGTAGA	(SEQ ID No. 80)
	5EJ	5'-AGATGTACTTTGGCCATTCCAG	(SEQ ID No. 81)
	5EK	5'-GCCATGACAGCAACATTATCTC	(SEQ ID No. 82)
	5EL	5'-CTTACTGCTACTGCAAGTTCTTC	(SEQ ID No. 83)
	5EM	5'-TCGATCAAAACAGTACAGGTG	(SEQ ID No. 84)
25	5EN	5'-GCAGATGTAGGAGACAAATCATC	(SEQ ID No. 85)
	5EO	5'-TCATCCAAAATCTCTAAATTTCTGG	(SEQ ID No. 86)
	5EP	5'-CTGAGGACCAGAACTGTATGC	(SEQ ID No. 87)
	5ES	5'-GCTGATTTGGTGTCTAGCCTGG	(SEQ ID No. 88)
	5ET	5'-TGCCTGGGTTGCAGGCCTGC	(SEQ ID No. 89)
30	5EX	5'-TTGGAAACAACCTGCATAGCAGC	(SEQ ID No. 90)

EXAMPLE 6

ISOLATION OF GENOMIC DNA CONTAINING WERNER'S SYNDROME GENE

5

To facilitate mutational analysis of the *WRN* gene, the intron-exon structure is determined. The *WRN* gene is located in the genomic sequence of P1 clone 2934. However, this clone only contains the 3' end of the gene (exons 21 to 35). Genomic clones containing the 5' end are obtained from a chromosome 8-specific cosmid library LA08NC01 (Wood et al. *Cytogenet. Cell Genet.* 59: 243, 1992) by
 10 screening for clones adjacent to P1 clone 2934. Briefly, this library is arrayed for PCR screening as described in Amemiya et al. (*Nucl. Acids Res.* 20: 2559, 1992). *WRN* containing cosmids are identified using primer sets 5E6/5EY, 5ED/5E12, and CD-A/CD-B (Table 3), which are derived from the *WRN* cDNA sequence (Figure 1;
 15 GenBank Accession No. L76937). Four walking steps yielded cosmids 193B5, 114D2, 78D8 and 194C3, which contained the remaining exons. Primers derived from the *WRN* cDNA were used for the initial sequence analysis of the cosmid clones. The resulting sequence (Figure 5) is compared to the cDNA sequence to identify intron-exon boundaries. Sequencing primers are then designed from the intron sequences to obtain
 20 sequence in the reverse direction and to obtain the second boundary defining the intron-exon junction. This strategy is used to define the exons not present in P1 clone 2934.

Table 3. Primer sequence and PCR conditions for *WRN* analysis

Region	Primer Sequence	Product Size (bp)	Mg ⁺² (mM)	pH
N-domain	5E6 5'-GATATTGTTTGTATTTACCCATGAAGAC (SEQ ID No. 164) 5EY 5'-TCCGCTGCTGTGCAGTTCTTTCC (SEQ ID No. 165)	106	1.5	8.3

The annealing temperature was 60° C for all primer sets.

Table 4 presents a summary of the structure of the genomic WRN gene.

- 5 The first column identifies the exon, the second column indicates the base numbers of the cDNA that are derived from the exon, the third column denotes the size of the exon in bp, the fourth column shows the sequence of the boundaries with intron sequences in lower case letters and exon sequences in upper case letters, the fifth column shows notable features of the exons.

10

Table 4. Intron-Exon Structure of the *WRN* Gene

Exon	cDNA Location	Exon Size (bp)	Intron-Exon Boundary Sequences	Exon Features
1	1-155	>155	...TTCTCGGGgtaaagtgtc (SEQ ID No. 169)	5' UTR
2	156-327	172	tacctctcagTTTTCTT...AAAGAAAGgtatgttgtt (SEQ ID No. 170)	5' UTR, ATG codon
3	328-440	113	taaactcaagGCATGTGT...GATATTAGgtaagtgatt (SEQ ID No. 171)	
4	441-586	146	ctcacttttagCATGAGTC...CATGTCAGgttggtatct (SEQ ID No. 172)	
5	587-735	149	aatgttacagTTTTCCC...ATAAAAAGgtaaaagcaa (SEQ ID No. 173)	
6	736-885	150	tcatttctagCTGAATG...ATGCTTAgttacgttctt (SEQ ID No. 174)	
7	886-955	70	ttttttataagGCTGGTTT...AAATAAGgtatgtttaa (SEQ ID No. 175)	
8	956-1070	115	ttccccctagAGGAAGAA...CCACGGAGgttaaataatt (SEQ ID No. 176)	
9	1071-1500	430	tttttttagGGTTTCTA...CTACTGAGgtactaaaat (SEQ ID No. 177)	
10	1501-	81	tttttaaaacATTATC...TGCTTAAGggtatgtttt (SEQ ID No. 178)	3' UTR, poly(A)
11	1503-1807	305	ttctttttagGCTTTAAG...GATTAAGgttgcattttt (SEQ ID No. 180)	

Exon	cDNA Location	Exon Size (bp)	Intron-Exon Boundary Sequences	Exon Features
13	1808-1883	76	ttattttcagACTTTTGG...TTTAAACCgtgagctataa (SEQ ID No. 181)	
14	1884-1951	68	caccttcaagAGTTCACT...GGCAACTGgtaagttgta (SEQ ID No. 182)	helicase motif I (5' end)
15	1952-2060	109	tcatttcaagGATATGGA...CAGCTTAAgtaagtcatg (SEQ ID No. 183)	helicase motif I (3' end) and Ia
16	2061-2129	69	cttccttatagAATGTCCA...ATTAAATTgtgagtaatt (SEQ ID No. 184)	
17	2130-2212	83	gtttttacagAGCTAAAT...TCATATTGgtaagtgata (SEQ ID No. 185)	
18	2213-2319	107	ttttttacagGTATCAGG...TGCCAATGgtaagctttg (SEQ ID No. 186)	helicase motif II
19	2320-2504	185	catcattcagGTTCCAAT...AAAACAAGgtaaggattt (SEQ ID No. 187)	helicase motif III
20	2505-2679	175	ttttcttttagTTCCCACT...AAATTCAGgtatgaggat (SEQ ID No. 188)	helicase motif IV
21	2680-2961	182	ttgtttctcagTGTGTGAT...TTAAATAGgtaaaaaaaa (SEQ ID No. 189)	helicase motifs V and VI
22	2962-2963	102	taatgcacagGCACCTTC...AGGAGACAgtatgtatta (SEQ ID No. 190)	
23	2964-3056	93	tcttgggtagAATCATCT...AGGTCCAGgtaaaagattt (SEQ ID No. 191)	
24	3057-3198	142	ttttatttagATTGGATC...GAGGATCTgtaagtatat (SEQ ID No. 192)	
25	3199-3369	171	ctaatttcagAATTCCTCA...CGAAAAAGgtaaacagtg (SEQ ID No. 193)	
26	3370-3464	95	cttttaattagGGTAGAAA...CTGCCTAGgtttaaatttt (SEQ ID No. 194)	
27	3465-3540	76	tttttttttagTTGAAAA...AGAAGAAGgtttgtttta (SEQ ID No. 195)	
28	3541-3614	74	ttaaatgcagTCTAACTT...AAAAAAGgtacagagtt (SEQ ID No. 196)	
29	3615-3690	76	aatatttttagTATCATGG...AGACTCAGgtaaaacattt (SEQ ID No. 197)	

FIG. 1. Schematic diagram of the cDNA sequence of the human DNA topoisomerase II α gene. The exons are numbered 1 through 29. The intron-exon boundary sequences are shown above the exons. The exon features are shown to the right of the exons.

Exon	cDNA Location	Exon Size (bp)	Intron-Exon Boundary Sequences	Exon Features
32	3919-4050	132	aattctgtagACAGACCT...TGCCCTTGgtaagtgtga (SEQ ID No. 200)	
33	4051-4213	163	ctttctctagAAGAGCAT...CAACTCAGgtgagaggca (SEQ ID No. 201)	
34	4214-4422	209	tcgtttacagATATGAGT...ATACTGAGgtattaatta (SEQ ID No. 202)	
35	4423-5190	768	tttctacagACTTCATC... (SEQ ID No. 203)	TAA codon, 3'UTR

Note. Exons are in uppercase and intron sequences are in lowercase letters.

As shown above, WRN contains a total of 35 exons ranging in size from 68 bp (exon 14) to 768 bp (exon 35). The coding region begins in the second exon (Table 2). As noted previously, there is a duplicated region in the WRN cDNA sequence which is 27 amino acids in length. This duplication is exactly conserved at the nucleotide level in cDNA. At the genomic level, the duplicated sequences were present as 2 exons (exons 10 and 11), each exon containing only the duplicated nucleotides. The intronic sequences adjacent to these 2 exons are also highly conserved, suggesting that the a relatively recent duplication event is responsible for these repeated exons. In addition, because the surrounding intronic sequences were conserved, it was not possible to design primers which could specifically amplify exons 10 and 11.

The helicase region of the WRN gene is contained in exons 14-21. Helicase motif 1 is split between exons 14 and 15 while the remaining motifs are each in an individual exon (Table 4). This region, from codon 569 to 859, has sequence similarity to the 7 signature helicase motifs. In addition, though the sequences between the motifs are not conserved, the spacing is very similar in genes from a wide range of species. For example, the helicase domains in the *E. coli RecQ* gene are found in a

EXAMPLE 7

IDENTIFICATION OF MUTATIONS

Initially, 4 different mutations in the C-terminal domain of *WRN* were identified. These mutations accounted for more than 80% of the Japanese WS patients examined. All 4 mutations are in the C-terminal domain region of *WRN* and the resulting predicted protein contained an intact helicase domain. Additional WS subjects are screened to identify further mutations. Genomic structure information is used to design PCR-primers for amplifying each exon, which is then subjected to DNA sequence analysis. Five additional *WRN* mutations are described; 2 are located in the consensus helicase motifs and another 2 are predicted to produce truncated proteins without the helicase domains. These mutations suggest that in at least some WS subjects, the enzymatic helicase activity is destroyed and support that complete loss-of-function of *WRN* gene product causes Werner's syndrome.

Although any cell may be used to isolate DNA, PBMC are preferred. As above, PBMC are obtained by venipuncture and subsequent hypotonic lysis of erythrocytes. PBMC are lysed by the addition of detergent, such as 0.5% NP-40, 0.5% Triton-X100, or 0.5% SDS. If a non-ionic detergent is used, no further purification of DNA is necessary, but proteinase K treatment, and subsequent heat killing of the enzyme (95°C for 10 minutes) is required. Genomic DNA is amplified according to the PCR conditions recited above using the primers listed in Table 5. Exons 9 and 10 are contained in a region of DNA that is duplicated. The primer pair for exon 9 and 10 anneals to sequences outside the duplication. Amplified product is analyzed by DNA sequence determination, hybridization with allele-specific probe, or other mutation detection method. When DNA sequences are determined, the sequence of the amplified exon is aligned with the known sequence (Figure 2A) and any discrepancies between patient samples and the reference sequence are identified.

Table 5

PCR Fragment	Primer Sequence	Product Size (bp)	Mg ²⁺ (mM)	pH
exon 1	A 5'-AGGGGCTCCACGCATGACGC (SEQ ID No. 92) B 5'-AGTCTGTTTTTCCAGAATCTCCC (SEQ ID No. 93)	583	1.5	8.3
exon 2	A 5'-CCTATGCTTGGACCTAGGTGTC (SEQ ID No. 94) B 5'-GAAGTTTACAAGTAACAACAGTACAC (SEQ ID No. 95)	339	1.5	8.3
exon 3	A 5'-ACTATAAATTGAATGCTTCAGTGAAC (SEQ ID No. 96) B 5'-GAACACACCTCACCTGTAAACTC (SEQ ID No. 97)	316	1.5	8.3
exon 4	E 5'-GGTAAACCACCATACCTGGCC (SEQ ID No. 98) F 5'-GTACATATCCTGGTCATTTAGCC (SEQ ID No. 99)	691	1.5	8.3
exon 5	B 5'-ATTCAGATAGAAAGTACATCTGTG (SEQ ID No. 100) E 5'-GTTAAGAAATACTCAAGGTCAATGTG (SEQ ID No. 101)	369	1.5	8.3
exon 6	A 5'-GGTTGTATTTTGGTATAACATTTCC (SEQ ID No. 102) B 5'-ATATTTTGGTAGAGTTTCTGCCAC (SEQ ID No. 103)	374	1.5	8.3
exon 7	A 5'-CTCTTCGATTTTCTGAAGATGGG (SEQ ID No. 104) B 5'-CCCTAATAGTCAGGAGTGTCAAG (SEQ ID No. 105)	291	1.5	8.3
exon 8	A 5'-GGAAAGAAAATGAAAATTTGATCCC (SEQ ID No. 106) B 5'-CAGCCTTAATGAATAGTATTCTTCAC (SEQ ID No. 107)	316	4.0	8.3
exon 9	C 5'-ATGATCTTTTAAAGTGAAGGTACG (SEQ ID No. 108) D 5'-CAAGTATCTCAAAAATTTGATC (SEQ ID No. 109)	668	1.5	8.3

PCR Fragment	Primer Sequence	Product Size (bp)	Mg ²⁺ (mM)	pH
	(SEQ ID No. 111)			
exon 13	A 5'-TAACCCATGGTAGCTGCACTG (SEQ ID No. 112) B 5'-CTGTTGCTGTTAAGCAGACAGG (SEQ ID No. 113)	285	1.5	8.3
exon 14	C 5'-TTGAATGGGACATTGSTCAAATGG (SEQ ID No. 114) D 5'-GTAGTTGCAITTTGTATTTGAGAGT (SEQ ID No. 115)	348	1.5	8.3
exon 15	E 5'-GTAAAAAGAAATGAAAGCATCAAAGG (SEQ ID No. 116) F 5'-TCACCCACAGAAGAAAAAGAGG (SEQ ID No. 117)	246	4.0	8.3
exon 16	A 5'-CAAAAAAGAAAAATTGCAAGAACAGG (SEQ ID No. 118) B 5'-CAGCAACATGTAATTCACCCACG (SEQ ID No. 119)	282	4.0	8.3
exon 17	S 5'-GAAGAGACTGGAATTGGGTTTGG (SEQ ID No. 120) S 5'-ATAGAGTATCATGGGATAAGATAGG (SEQ ID No. 121)	532	1.5	8.3
exon 18	A 5'-TTCTCCTTTGGAGATGTAGATGAG (SEQ ID No. 122) B 5'-TCTTCAGCTTCTTTACCAGTCCCA (SEQ ID No. 123)	273	4.0	10
exon 19	A 5'-CATGGTGTTTGACAACAGGATGG (SEQ ID No. 124) B 5'-GTTAAATATGCATTAGAAGGAAATCC (SEQ ID No. 125)	396	4.0	9.0
exon 20	A 5'-ATAAAACCAAACGGGTCTGAAGG (SEQ ID No. 126) B 5'-AAAAGAAGTATTCAATAAGATCTGG (SEQ ID No. 127)	342	4.0	8.3
exon 21	A 5'-AATTCACCTTTGTGCCAGGGACT (SEQ ID No. 128) B 5'-ACTTGGGATATTCGAAATAGCTT (SEQ ID No. 129)	397	1.5	9.0

PCR Fragment	Primer Sequence	Product Size (bp)	Mg ²⁺ (mM)	pH
exon 23	A 5'-CTGAAGTCAAATAATGAAGTCCCA (SEQ ID No. 132) B 5'-GTTTGCTTTCTGATATCTAAACACA (SEQ ID No. 133)	360	4.0	8.3
exon 24	A 5'-CTTGTGAGAGGCTATAAACTGG (SEQ ID No. 134) B 5'-GGTAAACAGTGTAGGAGTCTGC (SEQ ID No. 135)	267	1.5	8.3
exon 25	C 5'-GCTTGAAGGATGAGGCTCTGAG (SEQ ID No. 136) D 5'-TGTTGAGAATGAGCAGCATGGG (SEQ ID No. 137)	461	1.5	8.3
exon 26	A 5'-CTTGTGAGAGGCTATAAACTGG (SEQ ID No. 138) B 5'-GGTAAACAGTGTAGGAGTCTGC (SEQ ID No. 139)	267	1.5	8.3
exon 27	A 5'-GCCATTTTCTCTTTAATTGGAAAGG (SEQ ID No. 140) B 5'-ATCTTATTCATCTTTCTGAGAATGG (SEQ ID No. 141)	274	1.5	8.3
exon 28	A 5'-TGAAATAGCCCAACATCTGACAG (SEQ ID No. 142) B 5'-GATTAAATTTGACAGCTTGAATAGGC (SEQ ID No. 143)	291	1.5	8.3
exon 29	A 5'-TGAAATATAAACTCAGACTCTTAGC (SEQ ID No. 144) B 5'-GTACTGATTTGGAAAGACATCTC (SEQ ID No. 145)	303	1.5	8.3
exon 30	A 5'-GATGTGACAGTGGAGGCTATGG (SEQ ID No. 146) B 5'-GGAAAAATGTGGTATCTGAAGCTC (SEQ ID No. 147)	307	1.5	8.3
exon 31	A 5'-AAGTGAGCAAATGTTGCTTCTGG (SEQ ID No. 148) B 5'-TCATTAGGAACTGAACATCAGC (SEQ ID No. 149)	304	1.5	8.3

PCR Fragment	Primer Sequence	Product Size (bp)	Mg ²⁺ (mM)	pH
exon 33	A 5'-TAAAGGATTAATGCTGTTAACAGTG (SEQ ID No. 152) B 5'-TCACACTGAGCATTTACTACCTG (SEQ ID No. 153)	360	1.5	8.3
exon 34	C 5'-GCAAAGGAAATGTAGCACATAGAG (SEQ ID No. 154) D 5'-AGGCTATAGGCATTTGAAAGAGG (SEQ ID No. 155)	491	1.5	8.3
exon 35	A 5'-GTAGGCTCCAGAAAGACCCAG (SEQ ID No. 156) B 5'-GAAAGGATGGGTGTGTATTCAGG (SEQ ID No. 157)	406	1.5	8.3
mutation 7	GD A 5'-ACAGGCCATAGTTTGCCAACCC (SEQ ID No. 158) GD D 5'-TGGTATTAGAATTTCCCTTTCTTCC (SEQ ID No. 159)	426	1.5	9.0
DJG RT-PCR	SEE 5'-TGAAAGAGAATATGGAAAGAGGCTTG (SEQ ID No. 160) B 5'-CTTTATGAAGCCAAATTTCTACCC (SEQ ID No. 161)	2002	1.5	8.3
P2934AT1	A 5'-TCAAAATCAGTCGCCTCATCCC (SEQ ID No. 162) B 5'-CAATGTATCAGTCAGGGTTCACC (SEQ ID No. 163)	168	2.0	8.3

The annealing temperature was 60° C for all primer sets.

Mutations are detected by amplifying *WRN* exons from genomic DNA and directly cycle-sequencing the PCR products by dye-terminator cycle sequencing (Perkin Elmer) and an ABI373 automated DNA sequencer. Prior to sequencing, the PCR-amplified exon fragments were purified using a QIAquick 8 PCR purification kit (Qiagen). The resulting sequences are aligned by FASTA analysis (GCG). Nucleotide differences between WS and controls are subsequently confirmed by sequencing the reverse strand

consequences of splice-junction mutations. RT-PCR products were synthesized from

mRNA isolated from lymphoblastoid cell lines (Qiagen Oligotex, Qiagen). The large genomic deletion was detected in genomic DNA using long-range PCR (Expand Long Template PCR System, Boehringer Mannheim).

Diagnostic Criteria. WS patients were from an International Registry of Werner's Syndrome subjects. Diagnostic criteria are based on the following signs and symptoms (Nakura et al. 1994). Cardinal signs are: 1) bilateral cataracts; 2) characteristic dermatological pathology (tight skin, atrophic skin, pigmentary alterations, ulceration, hyperkeratosis, regional subcutaneous atrophy) and characteristic facies ("bird" facies); 3) short stature; 4) paternal consanguinity (3rd cousin or greater) or affected sibling; 5) premature greying and/or thinning of scalp hair; 6) positive 24-hour urinary hyaluronic acid test, when available). Further criteria are: 1) diabetes mellitus; 2) hypogonadism (secondary sexual underdevelopment, diminished fertility, testicular or ovarian atrophy); 3) osteoporosis; 4) osteosclerosis of distal phalanges of fingers and/or toes (X-ray diagnosis); 5) soft tissue calcification; 6) evidence of premature atherosclerosis (e.g. history of myocardial infarction); 7) mesenchymal neoplasms, rare neoplasms or multiple neoplasms; 8) voice changes (high pitched, squeaky or hoarse voice); 9) flat feet. Diagnostic classifications are as follows: "Definite", all cardinal signs (#6 when available) and any 2 others; "Probable", the first 3 cardinal signs and any 2 others; "Possible", either cataracts or dermatological alterations and any 4 others; "Excluded", onset of signs and symptoms before adolescence (except short stature since current data on pre-adolescent growth patterns is inadequate) or a negative hyaluronic acid test. Family designations are as previously used (Nakura et al. 1994; Goddard et al. 1996; Yu et al. 1996).

Mutations in WS Subjects. Initial screening of the WRN gene was based on sequence from only the 3' end of the gene (exons 23-35). Thus the first 4 mutations (designated 1-4, Table 3) were in the region 3' to the helicase domains. In this mutation screening, primers amplify exons 2-35 along with approximately 80 bp of

screened for mutations. These subjects were selected based on haplotype analysis that

suggested that each subject might have a different mutation (Yu et al. 1994; Goddard et al. 1996). A total of 30 Japanese and 36 Caucasian subjects were ultimately screened for each mutation by DNA sequence analysis of the appropriate exon.

5

Table 6. Summary of WRN Mutations

Mutation	Codon	Exon	Type of Mutation	Nucleotide Sequence	Comment	Predicted Protein Length
none						1432
1	1165	30	substitution	CAG (Gln) to <u>T</u> AG (terminator)	nonsense	1164
2	1305	33	substitution	CGA (Arg) to <u>T</u> GA (terminator)	nonsense	1034
3	1230	32	4 bp deletion	gtag- <u>AC</u> AG to gt-AG	4 bp deletion at splice-donor site	1247
4	1047-1078	24	substitution	tag-GGT to tag-GGT	substitution at splice-donor site	1060
5	369	9	substitution	CGA (Arg) to <u>T</u> GA (terminator)	nonsense	368
6	889	22	substitution	CGA (Arg) to <u>T</u> GA (terminator)	nonsense	888
7	759-816	20	substitution	CAG-gta to CAG-tta	substitution at splice-receptor site	760
8	389	9	1 bp deletion	<u>A</u> GAG (Arg) to GAG (Glu)	frame-shift	391
9	697-942	19-23	deletion (> 15 kb)	-	genomic deletion	1186

Table 7.
Mutation Status of WS Subjects¹

Mutation	Japanese WS Subjects		Non-Japanese WS Subjects	
	Homozygous	Heterozygous	Homozygous	Heterozygous
1	SY ^D			
2	HH ^D , HM ^D , MH ^M , NN ^D		GAR ^D	
3			SYR ^I	
4	FJ ^D , FUW ^D , HA ^I , HW ^D , IU ^D , JO1 ^D , JO2 ^D , KAKU ^P , KY ^D , MCI ^D , MIE2 ^I , SK ^D , ST ^D , TH ^I , TK ^M , TO ^D , ZM ^D , 78-85 ^I			
5	KO ^D , OW ^P	KUN ^I	EKL ^D , AG0780 ^I , AG4103 ^M	DJG ^P , CP3 ^I , NF ^M
6			CTA ^D	SUG1 ^P
7	WKH ^D			
8				FES ^I
9				DJG ^P , SUG1 ^P

- 5 ¹The diagnostic classification is as previously described (Nakura et al 1994).
Diagnosis categories: ^D Definite; ^P Probable; ^M Possible; ^I Insufficient data. The country of origin (ethnic
group) of non-Japanese subjects are: AG00780, USA (Caucasian); AG04103, USA (Caucasian); CTA,
England (India, East African, Asian); CP3, France (Caucasian); DJG, Germany (German); EKL,
Switzerland (German); FES, Germany (German); NF, France (Caucasian); SUG, USA (Caucasian); SYR,
10 Syria (Syrian). AG04103 and AG00780 were obtained as cell lines from the Aging Cell Repository
(Camden, New Jersey).

Five new WS mutations were detected in the WRN gene (designated 5-9,
Table 6). Two of the mutations (5 and 6) were single base substitutions creating
15 nonsense codons. Mutation 5 results in a C→T transition changing an Arg to a

Caucasian subjects were homozygous, and 1 Japanese and 4 Caucasians were
heterozygous for this mutation (Table 7). Mutation 6 is also a C→T transition changing

an Arg to a nonsense codon. One Caucasian WS subject was homozygous for this mutation, and a second was a compound heterozygote. The predicted protein product is 888 amino acids. A third substitution mutation (mutation 7) was a G→T change at a splice-receptor site, generating a truncated mRNA devoid of exon 20 and a prematurely
5 terminated WRN protein at amino acid 760. A single Japanese WS subject was homozygous for this mutation.

Two deletions were observed. One (mutation 8) is a 1 bp deletion at codon 389 resulting in a frame shift and a predicted truncated protein 391 amino acids long. This mutation is found in one Caucasian patient as a heterozygote. The second
10 (mutation 9) is a much larger deletion. This deletion was first observed in RT-PCR experiments when 2 different RT-PCR products were obtained from RNA prepared from subject DJG. RT-PCR products produced by primers 5EE and B (Table 5) yielded 2 different products, one with the expected size of 2009 bp, and a second, shorter product approximately 700 bp smaller. The DNA sequence of the shorter product
15 revealed that exons 19 through 23 were missing. To further establish the nature of this mutation, primers (exon 18A and exon 24A, Table 5) derived from the exons flanking this potential gross deletion (exons 18 and 24) were used to amplify genomic DNA from subject DJG using a long-range PCR protocol. A single 5 kb fragment was observed corresponding to the shorter RT-PCR product. (The normal fragment, which
20 is estimated to be > 20 kb was not observed.) The complete DNA sequence of this 5 kb fragment was determined and contained the expected 3' and 5' ends of exons 18 and 24, respectively. The exonic sequences were separated by intronic sequences adjacent to the 3' and 5' end of exons 18 and 24, respectively. No sequences from exons 19-23 were found in the 5 kb fragment. In other subjects and controls, the intronic sequence
25 in the intron 3' to exon 18 contained 531 bp of unique sequence followed by a 241 bp Alu repeat element. Likewise, for the region 5' to exon 24, there is an Alu repeat element separated from exon 24 by 3,460 bp of unique sequence. The 4938 bp

recombination error at 2 highly homologous Alu elements within the WRN gene. A

primer set, GD-A and GD-D (Table 5) was designed to specifically amplify a short fragment (426 bp) across this junction point. A single additional Caucasian WS patient, SUG, was shown to contain this genomic deletion. Further PCR amplification of the exons within this deleted region demonstrated that both DJG and SUG are heterozygous for this mutation.

Origins of WRN Mutations. Because multiple subjects have the same mutation and because the same mutation was observed in different ethnic groups, at least some of the mutations likely originated in common founders. Evidence for a common founder was examined using 2 short tandem repeat polymorphisms (STRPs) within the WRN gene. These STRPs, D8S2162 and p2934AT1, were isolated from the same P1 clone (p2934) and are within 17.5 kb of each other. While D8S2162 is not particularly polymorphic (heterozygosity = 54% in Japanese and 70% in Caucasians) and is primarily a 2 allele system (140 and 142 bp alleles), p2934AT1 is highly polymorphic (heterozygosity = 78% in both Japanese and Caucasian populations). For mutation 4, which has only been observed in Japanese subjects, all but 1 subject had the D8S2164/p2934AT1 haplotype of 140-148 (Table 8). The single exception, JO2, has the haplotype 140-150, with the p2934AT1 allele being 2 bp different from the 148 bp allele observed in other subjects with mutation 4. This 2 bp difference may be the result of a 2bp mutation, as is commonly observed in dinucleotide repeat STRP loci (Weber and Wong, 1993). The haplotype data is consistent with a common Japanese founder and is consistent with the linkage disequilibrium observed in the same Japanese subjects for other markers in the WRN region (Yu et al. 1994; Goddard et al., 1996). For mutations 2 and 5, in the Japanese, the 896R18-p2934AT1 haplotypes for the small number of available subjects, are consistent with common founders for each mutation. However, the non-Japanese subjects with mutations 2 and 5 have discordant p2934AT1 genotypes when compared to Japanese subjects with the same mutations. These results do not support a common founder for both Japanese and non-Japanese subjects with

discordant for p2934AT1 (e.g. compare AG00780 to FK1). It should be noted that

absence of evidence for a common founder does not necessarily exclude the possibility of a single originating mutational event. Intragenic recombination and/or mutations creating new alleles at the 2 STRP loci could, over time, obscure the origins of the different WRN mutations.

Table 8. STRP Genotypes at the WRN gene¹.

Subject	Ethnic Group	Mutation	y896r18	p2934at1
FJ, FUW, HA, HW, JO1, KAKU, KY, MIE2, TO	Japanese	4	140/140	148/148
JO2	Japanese	4	140/140	150/150
HM, MH, NN,	Japanese	2	140/140	144/144
GAR	Hispanic	2	140/140	156/156
OW, KO	Japanese	5	140/140	148/148
AG00780	Caucasian	5	142/142	136/136
EKL, AG04103	Caucasian	5	142/142	128/128
CP3	Caucasian	5/?	142/150	128/142
KUN	Japanese	5/?	140/142	128/148
DJG	Caucasian	5/9	140/142	128/del ²

¹Genotype data for HH, SK, ST, TH, TK, and ZM was not available. For y896R18, alleles in bp (frequency for Caucasians, frequency for Japanese) were as follows: 136 (0.030, 0.025); 138 (0.020, 0.010); 140 (0.460, 0.576); 142 (0.337, 0.359); 144 (0.084, 0.010); 146 (0, 0.010); 148 (0.009, 0.010); 150 (0.059, 0). For p2934AT1, alleles in bp (Caucasian frequency, Japanese frequency) were as follows: 114 (0.006, 0); 122 (0, 0.009); 124 (0.011, 0); 128 (0.253, 0.079); 130 (0, 0.018); 132 (0.006, 0.009); 134 (0.046, 0.096); 136 (0.086, 0.009); 138 (0.011, 0); 140 (0.034, 0); 142 (0.052, 0.035); 144 (0.023, 0.061); 146 (0.023, 0.053); 148 (0.034, 0.132); 150 (0.034, 0.105); 152 (0.057, 0.123); 154 (0.063, 0.088); 156 (0.086, 0.070); 158 (0.098, 0.070); 160 (0.046, 0.018); 162 (0.029, 0.009); 166 (0, 0.009); 168 (0, 0.009).

The 5 mutations identified here demonstrate that WS mutations are not restricted to the 3' end of the gene, but are also found in other regions of WRN. In addition, mutations 5 and 7-9 each disrupt either part or all of the helicase region. Thus the WS subjects homozygous for this mutation will completely lack the WRN helicase

properly results in complete loss of all activity of the WRN protein. However, the WS phenotype in these subjects is not appreciably distinct from the WS phenotype

generated by the other mutations described here. Thus, all mutations in the WS gene may be complete loss of function mutations.

EXAMPLE 8

IDENTIFICATION OF MOUSE WRN GENE

The mouse WRN cDNA was isolated by screening a mouse splenocyte cDNA library at low stringency with human WRN cDNA as probe. The mouse cDNA sequence is presented in Figure 9. The homology between human and mouse WRN cDNA sequence is about 80%. On the amino acid level, the human and mouse WRN gene product show about 90% identity. Notably, the repeated exon in human WRN cDNA (exons 10 and 11) is only present once in mouse WRN cDNA.

Genomic mouse WRN clone was isolated by using mouse WRN specific primers to screen mouse genomic BAC library. The genomic DNA sequence is presented in Figure 6.

The genomic DNA sequence is presented in Figure 7 and SEQ ID NOS: 207-209. The DNA sequence is presented in Figure 6 and SEQ ID NOS: 205 and 206.

EXAMPLE 9

LOCALIZATION OF THE WRN GENE PRODUCT

A rabbit polyclonal antiserum raised to a peptide of WRN gene product is used in an indirect immunofluorescence assay to determine the intracellular localization of the WRN protein.

A rabbit polyclonal antiserum is raised to the peptide Phe-Pro-Gly-Ser-Glu-Glu-Ile-Cys-Ser-Ser-Ser-Lys-Arg (FPGSEEICSSSKR) (SEQ ID NO: 204) by standard methods (see Harlow and Lane, *Antibodies, A Laboratory Manual*, CSH Press, Cold Spring Harbor, 1989; *Current Protocols in Immunology*, Greene Publishing, 1995). The peptide corresponds to residues 1375 through 1387 of the WRN

fixed with 3% paraformaldehyde and permeabilized for 2 min with a buffer containing

0.5% Triton X-100, 10 mM PIPES, pH 6.8, 50 mM NaCl, 300 mM sucrose, and 3 mM $MgCl_2$ (see for example, Fey et al., *J. Biol. Chem.* 98: 1973, 1984). The cells are then stained for 20 min with a suitable dilution of the anti-peptide antibody (1:1500), washed, stained with a suitable second antibody (e.g., FITC-conjugated goat anti-rabbit antibody), washed, and mounted for visualization by fluorescence microscopy. Control stains include bis-benzimidine (Sigma, St. Louis, MO), which stains DNA, and phalloidin (Molecular Probes, OR, BODIPY 558/568 phalloidin), which stains filamentous actin.

As seen in Figure 9, the WRN gene product is almost entirely located in the nucleus. Nuclear staining is readily noted in the epithelial cells at the bottom left in panel A. These cells are close to the periphery of the expanding clone of human prostate epithelial cells. Cells that are not rapidly dividing (e.g., cells closer to the center of the clone), such as those seen in the upper right of panel A, are stained in both the cytoplasm and nucleus. The location and size of the nuclei in these cells is shown by staining DNA with the intercalating dye bis-benzimidine (Hoeschst 33258), panel B. The overall size of the cells and in some cases key cytoskeletal features are revealed by staining for F-actin as shown in panel C.

EXAMPLE 10

ISOLATION OF A PROTEIN THAT BINDS TO THE WRN GENE PRODUCT

A yeast 2-hybrid interaction screen (Hollenberg et al., *Mol. Cell Biol.* 13: 3813, 1995) is used to identify and isolate a cellular protein that binds to the carboxy-terminal 443 amino acids (residues 990 through 1432) of the WRN gene product.

A library of 1.1×10^6 independent cDNA clones generated from RNA isolated from stimulated human peripheral blood mononuclear cells is generated in pACT-2 (Clontech, Palo Alto, CA) that creates cDNA/GAL4 activation domain fusions is co-transfected into yeast containing pLEXA with the WRN gene fragment to generate

for histidine. Of these, 60 were cured of the pLEXA

plasmid by growth on medium containing cycloheximide and mated with a yeast strain expressing a fusion of a "sticky" laminin and the GAL4 activation domain. 19 clones did not activate the sticky protein and underwent DNA sequence analysis. Of these, 6 contained sequences that did not match any sequence in GenBank by BLAST search.

- 5 Two other clones encoded carnitine palmitoyl transferase I and prolyl 4-hydroxylase B subunit. Six independent clones encoded a 70K component of the U1 snRNP complex (GenBank Accession No. M22636). Moreover, all six derived from the RNA recognition motif region of the 70K protein.

- 10 From the foregoing, it will be appreciated that, although specific embodiments of this invention have been described herein for the purposes of illustration, various modifications may be made without departing from the spirit and scope of the invention. Accordingly, the invention is not limited except by the appended claims.

Claims

We claim:

1. An isolated nucleic acid molecule encoding a WRN gene product.
2. An isolated nucleic acid molecule selected from the group consisting of:
 - (a) an isolated nucleic acid molecule as set forth in the Figures or complementary sequence thereof;
 - (b) an isolated nucleic acid molecule that specifically hybridizes to the nucleic acid molecule of (a) under conditions of high stringency; and
 - (c) an isolated nucleic acid that encodes a WRN gene product.
3. An expression vector, comprising a promoter operably linked to a nucleic acid molecule according to any one of claims 1 or 2.
4. The expression vector according to claim 3 wherein said promoter is selected from the group consisting of CMV I-E promoter, SV40 early promoter and MuLV LTR.
5. The expression vector according to claim 3 wherein said promoter is a tissue-specific promoter.
6. A viral vector capable of directing the expression of a nucleic acid molecule according to claims 1 or 2.
7. The viral vector according to claim 6 wherein said vector is selected

8. A host cell carrying a vector according to any one of claims 3 to 7.
9. The host cell according to claim 8 wherein said cell is selected from the group consisting of human cell, dog cell, monkey cell, rat cell and mouse cell.
10. An isolated protein comprising a WRN gene product.
11. An antibody which specifically binds to the protein according to claim 10.
12. The antibody according to claim 11 wherein said antibody is a monoclonal antibody.
13. The antibody according to claim 11 wherein said antibody is selected from the group consisting of an Fab fragment, an Fv fragment and a single chain antibody.
14. A hybridoma capable of producing an antibody according to claim 12.
15. A nucleic acid probe which is capable of specifically hybridizing to a WRN gene under conditions of high stringency.
16. A pair of primers capable of specifically amplifying all or a portion of a nucleic acid molecule according to any one claims 1 or 2.
17. A transgenic animal whose germ cells and somatic cells contain a WRN gene which is operably linked to a promoter effective for the expression of said gene, said gene being introduced into said mouse, or an ancestor of said mouse, at an embryonic

18. The transgenic animal according to claim 17 wherein the animal is selected from the group consisting of a mouse, a rat and a dog.

19. The transgenic animal according to claim 17 wherein WRN is expressed from a vector according to any one of claims 3 to 7.

20. An agonist of a WRN gene product.

21. An antagonist of a WRN gene product.

FIGURE 1

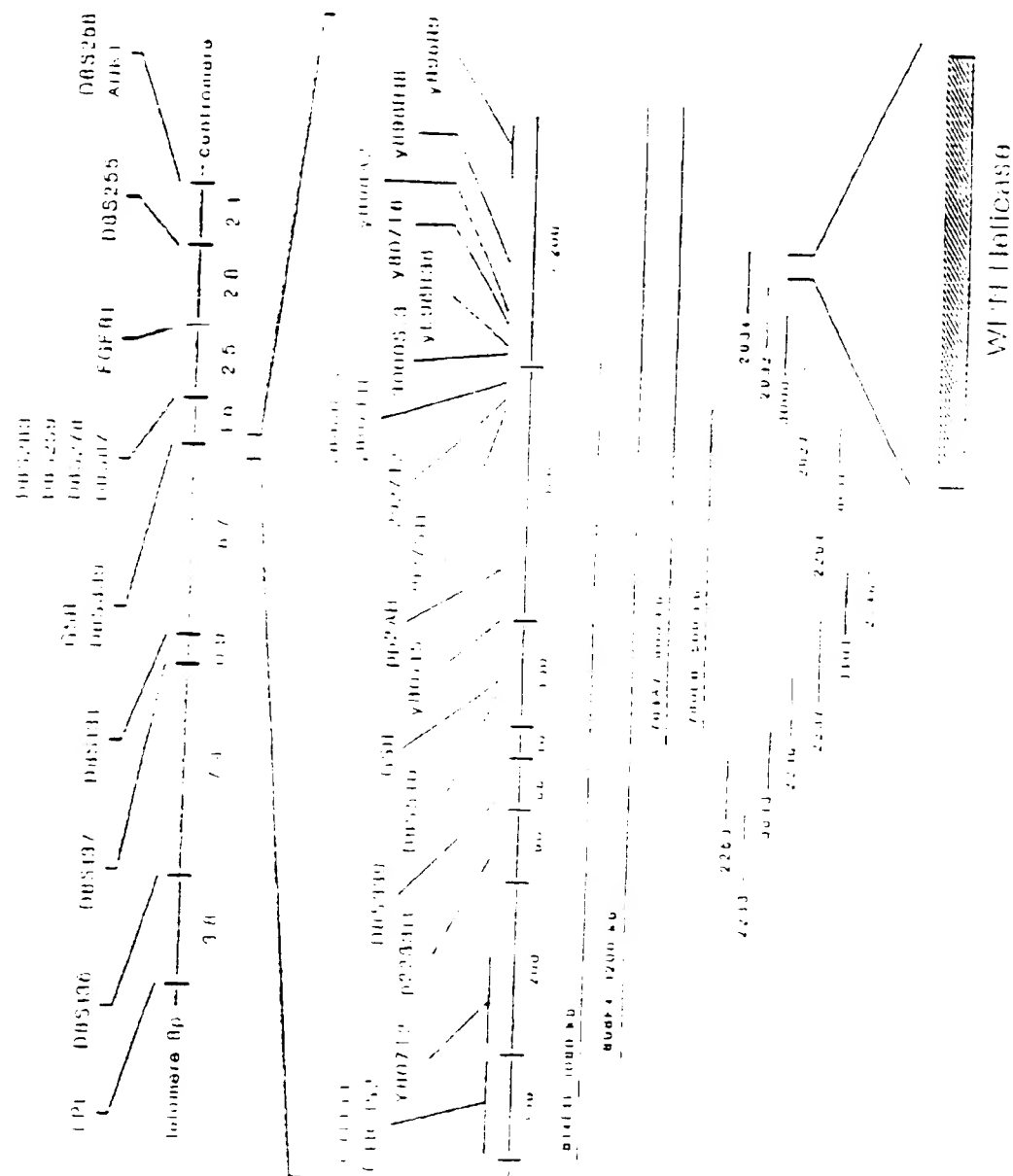


FIGURE 2A

[illegible]

1. *Prüfung* 2. *Prüfung* 3. *Prüfung* 4. *Prüfung* 5. *Prüfung* 6. *Prüfung* 7. *Prüfung* 8. *Prüfung* 9. *Prüfung* 10. *Prüfung* 11. *Prüfung* 12. *Prüfung* 13. *Prüfung* 14. *Prüfung* 15. *Prüfung* 16. *Prüfung* 17. *Prüfung* 18. *Prüfung* 19. *Prüfung* 20. *Prüfung* 21. *Prüfung* 22. *Prüfung* 23. *Prüfung* 24. *Prüfung* 25. *Prüfung* 26. *Prüfung* 27. *Prüfung* 28. *Prüfung* 29. *Prüfung* 30. *Prüfung* 31. *Prüfung* 32. *Prüfung* 33. *Prüfung* 34. *Prüfung* 35. *Prüfung* 36. *Prüfung* 37. *Prüfung* 38. *Prüfung* 39. *Prüfung* 40. *Prüfung* 41. *Prüfung* 42. *Prüfung* 43. *Prüfung* 44. *Prüfung* 45. *Prüfung* 46. *Prüfung* 47. *Prüfung* 48. *Prüfung* 49. *Prüfung* 50. *Prüfung* 51. *Prüfung* 52. *Prüfung* 53. *Prüfung* 54. *Prüfung* 55. *Prüfung* 56. *Prüfung* 57. *Prüfung* 58. *Prüfung* 59. *Prüfung* 60. *Prüfung* 61. *Prüfung* 62. *Prüfung* 63. *Prüfung* 64. *Prüfung* 65. *Prüfung* 66. *Prüfung* 67. *Prüfung* 68. *Prüfung* 69. *Prüfung* 70. *Prüfung* 71. *Prüfung* 72. *Prüfung* 73. *Prüfung* 74. *Prüfung* 75. *Prüfung* 76. *Prüfung* 77. *Prüfung* 78. *Prüfung* 79. *Prüfung* 80. *Prüfung* 81. *Prüfung* 82. *Prüfung* 83. *Prüfung* 84. *Prüfung* 85. *Prüfung* 86. *Prüfung* 87. *Prüfung* 88. *Prüfung* 89. *Prüfung* 90. *Prüfung* 91. *Prüfung* 92. *Prüfung* 93. *Prüfung* 94. *Prüfung* 95. *Prüfung* 96. *Prüfung* 97. *Prüfung* 98. *Prüfung* 99. *Prüfung* 100. *Prüfung*

FIGURE 2A (CONT.)

TTACTTTCTCCGAGTATCTAATTCTCAGGCTTTTCCGATCATATCCGAGGCGACAGTTTA	3240
TTTGGGACTGGGAGGRTCAACAGAGAGACTTGGTGGAGGCTTTTTCCTTTCAGCTGATC	3300
ACTGAGGGATTTGTTGGTAGAAGTTTCTGGGTATACCAATTTATGAAGATTTTGGGCGCTT	3360
ACGAAAAAGGGTAGAATTTGGCTTCATAPAGCTAATACAGAACTTCAGAGGCTCATGCTT	3420
CAAGCTAATGAAGATTTCTGTCCAAAGAGATTTTCTTCTGCTAGTTTCGAAACTGTATCT	3480
TGGGGGACCCAAAGAGATTTGTTATATCTAGTACCACTTGAATTAAGTACAGAGAGAGAG	3520
TCTAACTTTGAGAGATTTATATTTCTTATTAACCATCTGATTAAGATTTCTTCTGGGAGTAAC	3600
ATTTCTAAAAAAGTATCATGGTAGAGTCCACCGAAGAAAGCTTACAGTTCTTGAAGGCTT	3660
GTTATTTTGGGACAGAGAGAGGAGACTCAGATTTCTGTTACATGGCAAACTTGTAGAGCT	3720
AGGCGAAGACATGCCATATAAATGGATGTTCCCGGAGCTATTCTGGCAACAAACAGATA	3780
CTGGTGGATATGGGCAAAATGAGAGCCACTACGGTTGAAAGCTAAAAAGGATTCATGGT	3820
GTTTTCTGAAGGCAAGCTGCCATGTTGGGCGCTTCTGTTGGAGTCACTCAAAAGATTTCTGC	3900
CAAAACAAATAGTGTTCAGACAGAGCTCTTTTCAATACAAACCTCAAGAAAGAAAGAAAG	3960
AGGAGTGTGTTAGCAAAAAATAAAAATATGCTCAATTTTCAGCTCTATGGGCACTCAAGAAC	4020
TCTTTATTCGAGAGAAAGAGATTTCTTTTGAAGAGCATATCTGAGAGCAAGCACTCTGCT	4080
CTCATGCAATTTGGATGCACTTATCTGAAAGGCTGAAAGCTGCTGCTGCTGCTGCTGCTG	4140
GAGCTGAGAGGCTTCACTGAGAGCTTCAAGAGATTAATGCTGATGTTATGGGAAAGCTT	4200
CGGCTCACTCAAGCTTCACTTAATAATTAAGCTAATCAAGATCTTACTTTCTGAAAGATTT	4260
GAGAGCTAGCTTATGCAATGGCAATGAGATCTTAAAGATGCTCTGCAAGGGGAGCTT	4320
CAAGCTTCACTGCTGATCTCAAGAAAGAGAGATTTCTGCGGCTTCTGAAGATCTGCTTCA	4380
AGTTCTAAG	4440
AG	4500
AAAAAG	4560
TGGTCTATTATAG	4620
CTTAAG	4680
GGCTTCCGCAATTCATCTACTTTCTGGGTCTTCTGGGAGGCTAGCTGATCTACTACTCTA	4740
AGAGAGATTAATAAGTAGAGCTTCTGCTAGAGATTTCTTTAAGAGAGCTTTACTGCTGCTT	4800
TCTAATCTCTTTATTAAG	4860
ATTAAG	4920
TTCTGTTTTCTAAG	4980
AAG	5040
CGCATTTTGGGTGCAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	5100
TTGCACTAGGAG	5160
TAAAG	5208

FIGURE 2B

MSEKMLTTTAQQRKQFENMNTQNKRCAYEERKQVRAKSTFEDDLPTLEPTGSDITYSYDAS	60
DCSFLSEDIISMSLSDGGVVGTEMENPPLYNRGLGKVALDCLWSESNOYLPHYSSMSVF	120
PQGLHMLLENKRWKKGAGVGLSDQWKLRLRFDTHLKHIVELTDVANKKIKOTETWSLNSL	180
VKHLGKQLLNDKSEIRCSNWSKEPLTEDQKLYRATORAYAGFTIYFNLEILDSTVQRFAIN	240
KEEEILLSDMOKQLTGISSEVMDLAKHLPHAFSKLENFRAYSILLKDIENLYSLRAMII	300
GSTNIEHELRLPSNNLNLSSPEDSTTGGVQKQIRSEHWLEHWEDSTWQPTLDHLAKHDGE	360
DVLGNHVERKEDGGFEDGVEDNKLKENMERACLMSSLDITTEHELQLEQQSQSEVYLSDIAYK	420
STEHLSPNDNENDTSYVIESDEDELEMEMLFWLSPNDNENOTEVYVIESDEDELEMEMLKSLE	480
NLNSGTVEPTHSKCLKMERNLGLPTKEEEDDEENRNEGEEDDCKDFLNRPAPNEEQVTCL	520
KQVFGHSSFKPVQWVTHSVLEERRDNVAVMATGVYKSLQFQVFPVYVGNIGLYVDSPLIS	580
LMEDQVLRKMSNIPACTLGSAGSENVLTQIKLGNVYVYVTPYCSGIMGLLQQLERDI	600
GITLLIAYDERHCHSENGHDFRDSERKLGELKTALPMVPTVALTATASSSTRCDIVRCNL	660
RNPQTCTGFDPRNLYLEVRAKIGNILQDLQPLVNTSSHWTEGFTIDYOPSRMOTQQV	720
TGELRNTNLSDCTYKAGMSTETAKCITHRFVRDETCQVIATLAPENGINKVDIRQVINY	780
AFKIMESVYQELGRAGROGLSSCHVILWAPADINLRAHLLTIRVINTPLNMLPDMANIE	840
KILHSEFPQFQILLRTEEDKQVTHRELDMOTENCOONOFSLRCHVEMDSEEDTSWDFG	900
PQAEVLLSANDILGENTGIELLFLRGENSLPLADQVSEHELQOTONDQTESWVWKRTER	960
QLITEEFLEWERLNNKTHICALTHNGPNNLHWNTESLRLILQVTELDKXHTLPSKH	1020
TVSSGTHVHONVQVVELSTEWSENLLVLEWVWFOCNLSSGENTSYVYDNTQSPENKUSE	1080
SGPVCSAIGQETQVLYVHLYEAFCKHANNHNEFPALLADYHILVDMANMRPTTVENTHR	1140
LDGVSEGNLAPLAPLEVTXNFTQNMVATCLFESTWLEEDQWTELWANDHICTLSQEMV	1200
ITVSLFTEKXKPLKSLAESRILPLNTIGMHLSTAVYAGCPLEDRAGLTPEVQKILRDVY	1260
RNEFPWSEDMSKLILPMLVPEMIDTULIHMATECLNNGFSSGLQPSODVNRACFPSSSE	1320
LCSSGHRASKEEINGINTETSSAEKARALPITWANGESTENGLMDKTKRGLFS	1380
	1422

FIGURE 3A

[illegible]

FIGURE 3B

CTTGAGAGGACAGAGAGACGCTTTCTTACGAAAAATATAATATTCACGTTTTCACAG 972
 TTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1000
 TTTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1032
 SerMetAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 240
 GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1092
 GluSerArgGluLeuProLeuMetThrAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 250
 GCGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1152
 GluGlyProLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 1200
 GATTTATTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1212
 AspValAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 1220
 TTTATTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1272
 LeuValProLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 1300
 GGTGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1312
 GluProAspSerGluLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 1340
 TTTATTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1392
 SerLeuAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 1400
 ACTTTGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1432
 ThrSerSerLeuAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 1440
 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1492
 SerGlyGlyLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 1500
 CAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1572
 TTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1612
 AACTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1632
 AGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1672
 ACTGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1712
 TCGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1732

FIGURE 3C

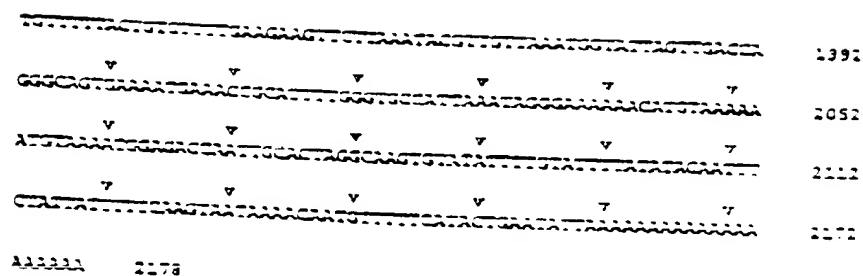


FIGURE 1A

27 heliceses mat (Agem_12_27_12.pro)	1	50
n 12_27 heliceses mat (recg_ecoli.pro)		
n 12_27 heliceses mat (YAbc_SCHPO.pro)		
n 12_27 heliceses mat (recg_human.pro)		
agpm_12_27 heliceses mat (HLM.pro)		
MAAGVQHHQ LQDLEHGLT LQDLEEP KESGTEKK ESSDHVSVT		
37	100	
27 heliceses mat (Agem_12_27_12.pro)		
n 12_27 heliceses mat (recg_ecoli.pro)		
n 12_27 heliceses mat (YAbc_SCHPO.pro)		
n 12_27 heliceses mat (recg_human.pro)		
agpm_12_27 heliceses mat (HLM.pro)		
HWAVAEIVL RHLDVNLID LQDLEPLD TRQQRVKDF KNAPAGQETQ		
MTVETRL		
101	150	
27 heliceses mat (Agem_12_27_12.pro)		
n 12_27 heliceses mat (recg_ecoli.pro)		
n 12_27 heliceses mat (YAbc_SCHPO.pro)		
n 12_27 heliceses mat (recg_human.pro)		
agpm_12_27 heliceses mat (HLM.pro)		
IRHLDHFRS SQQLDIVER PLELDLVRV KVSSDIVVK DSDPKSRV		
RDGKSLDQ LQGLPLVA LQGLTVRK SDEALKKE FSSSEHSLST		
151	200	
27 heliceses mat (Agem_12_27_12.pro)		
n 12_27 heliceses mat (recg_ecoli.pro)		
n 12_27 heliceses mat (YAbc_SCHPO.pro)		
n 12_27 heliceses mat (recg_human.pro)		
agpm_12_27 heliceses mat (HLM.pro)		
FDDEDAVAT LLEPLDGS		
SSNNLK HKDVEGPHLL		
THADPHDQ LQGLSKAV TRQQLVRV SVAQSKKGK RNFKAQIYT		
201	250	
27 heliceses mat (Agem_12_27_12.pro)		
n 12_27 heliceses mat (recg_ecoli.pro)		
n 12_27 heliceses mat (YAbc_SCHPO.pro)		
n 12_27 heliceses mat (recg_human.pro)		
agpm_12_27 heliceses mat (HLM.pro)		
KPKKLAIPA LQSDHLDLV DLDLKAQM EVFQIQPLA VNFADTVSH		
THLVETDAP KESGQQLT LQGLDGLR LQSDVCTDD GPFAEVITRE		

FIGURE 4D

851	800
50	YCSNMGLQ QLEADIGTID
51	REHEDRL EHLAIDHIVL
52	SRGATRVRLK SLYRKLAR
53	SGEMMSRLK KAYLAKETIR
54	ASDRLEETLE HLYRKLAR
851	850
55	PMVPIVALTA TASSSTIKRI
56	ETLPTMALTA TADDETIQRI
57	GLPTMALTA TARELIVAKNI
58	THASLIGITA TAKHIVLITIA
59	PSVPIVMALTA TAREIVQRI
851	900
60	NILOQLQPEL VKTSSHWEP
61	KPLAQIM RYVQEGRKKS
62	KDLYPELY RFLNGHILHE
63	RIEDEFEDIV KETHGRYKQ
64	IKL LWRKHHYD
901	950
65	YUAGMSFSTR KDIHIREVR
66	YUAGLEHIVR ADVQEKQIR
67	YUAGLEKVER QRIQHEW QS
68	YUAGLEPEEK ITVIRKWA
69	YUAGLSISAR DEVQQRIRIQ
951	1000
70	KOMESYYQET CHAGRIRKIQS
71	RRIESYYQET CHAGRIRKIPA
72	RSLEGYYQET CHAGRIRKTA
73	KSHIHYYQES CHAGRIRMEA
74	NAVIGYYQES CHAGRIRKLS

FIGURE 4B

27 helices mat (Agaric 12 27 f2 pro)	1251	ESLHLLKLYS YHSDHLLSG QHDSKSTHFV QSDPKAYSS QVLSAQDQ	1400
12 27 helices mat (veg ecoli pro)	1401	ELHAPDK	
12 27 helices mat (YAB SCHW pro)		YHSDHLLSG QHDSKSTHFV QSDPKAYSS QVLSAQDQ	
12 27 helices mat (veg human pro)		ESLHLLKLYS YHSDHLLSG QHDSKSTHFV QSDPKAYSS QVLSAQDQ	
Agaric 12 27 helices mat (hlt pro)		ESLHLLKLYS YHSDHLLSG QHDSKSTHFV QSDPKAYSS QVLSAQDQ	
27 helices mat (Agaric 12 27 f2 pro)	1401	ESLHLLKLYS YHSDHLLSG QHDSKSTHFV QSDPKAYSS QVLSAQDQ	1450
12 27 helices mat (veg ecoli pro)		ELHAPDK	
12 27 helices mat (YAB SCHW pro)		YHSDHLLSG QHDSKSTHFV QSDPKAYSS QVLSAQDQ	
12 27 helices mat (veg human pro)		ESLHLLKLYS YHSDHLLSG QHDSKSTHFV QSDPKAYSS QVLSAQDQ	
Agaric 12 27 helices mat (hlt pro)		ESLHLLKLYS YHSDHLLSG QHDSKSTHFV QSDPKAYSS QVLSAQDQ	
27 helices mat (Agaric 12 27 f2 pro)	1451	ESLHLLKLYS YHSDHLLSG QHDSKSTHFV QSDPKAYSS QVLSAQDQ	1500
12 27 helices mat (veg ecoli pro)		ELHAPDK	
12 27 helices mat (YAB SCHW pro)		YHSDHLLSG QHDSKSTHFV QSDPKAYSS QVLSAQDQ	
12 27 helices mat (veg human pro)		ESLHLLKLYS YHSDHLLSG QHDSKSTHFV QSDPKAYSS QVLSAQDQ	
Agaric 12 27 helices mat (hlt pro)		ESLHLLKLYS YHSDHLLSG QHDSKSTHFV QSDPKAYSS QVLSAQDQ	

FIGURE 5B

ATTACAGAG	TGTTGATGAA	TATGCTGCG	ATGGACTTGA	AAAAACAAAC	AGCTGCTTGA	4250
TGTTGATGAA	GAATTTGAA	TAAATTTGAT	ATGGCAAAAC	AATGCTTAT	NATGATTTGA	4300
GAAAGGAACT	TAAATTTGAT	AGTTGACTGG	TGTGCBATAA	TAACGCTTAA	TACTAAGAGA	4350
CACTTTTGA	GGGCGAGAG	TGCTGAGAG	CTGATAGAGA	TAAAGCTTGA	NATTTGAGCT	4400
TGTTGATGAA	TGTTGATGCT	GAATGCTTGA	AGTTTCTCAT	ACAAATATAA	ATTAAGAGATA	4500
AGCTTTTGA	TGGGAAAGG	TAAATTTGA	TAAATTTGAG	AAGCAATTTT	TGAAGGCAAA	4550
AACTTTGCTGA	GAATTAATTT	GAATTAATTT	AGCTTAGAGT	TTTTCATAGT	TTTTTGAAGC	4600
AGCTTAAGAA	TGATAGCTTG	GACACAGGAA	GAATTTTCTT	TGAAGAGCAA	TTTTTATATA	4650
ATGCAAGAAA	TGTTGATGCT	TTGTTGGGGG	GATTTGACTG	AAAGGGAATC	NACAGAACTT	4700
TGTTGATGCT	TAGAAATGCT	TTTATATGCT	GATGGGGGCT	GGGTTTCTGA	GATAATGAAA	4800
AATAACAGCT	AAAAATTAAG	TAAAAAATAA	AGTAAGAGAG	TTGCGAATAC	AGTTTTCAT	4850
ATGCTTTGCT	TGTTGATGCT	GGACAGGAGC	CTTCTTACTG	AGATAGCTTA	TGAGAACTTT	4900
CGATTTATCA	AATTAAGAT	GATGGCAAGC	ATGGAAAAAT	ATTTTCTATC	TAGCAGATTT	4950
AGGAGAGAGT	ATTTATATTT	TATTTTATTC	CAATAGTTATG	GAATTTATGA	TGATGCTTTT	5000
TTAAGAGAGC	AAATTTGCTTA	AAATTTATTC	AGTAATTTGA	AAAAATATTT	TGTTGCTGCT	5100
AGATTTGCTG	TGAATTTTGA	AGGCTAACTT	CTTTGCTGCT	GAATTAATTC	AACTTAATA	5150
ATGAAAGTGG	AACTTAATTC	ATTAATGCTG	ATTTTCTGCT	ATTAATTTGA	GGAACTGAAT	5200
TAACATATAC	CTTTGCTGCT	TTTTATTTTA	TTTTAAAAAT	TGTTGCTGCT	AGAACTGCTT	5250
TGTTGCTGCT	TGAGGAGCAA	GAATTTATTC	AACTTTGCTT	GGAAATTTAT	GGAACTGAAA	5300
AACTTTGCTT	TAAATTTGAA	TGAGGCTAAA	GAATTTGCTT	TATAGATGGA	GAATTTAAAA	5350
GTGCTTTGCT	CTTTGCTGCT	TGATTTGCTT	GTGAAATTTT	GTGAAATTTT	AACTTTATTC	5400
GTGCTTTGCT	GAATTTGCTT	AACTTTGCTT	GAATTTGCTT	AACTTTGCTT	AACTTTGCTT	5450
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	5500
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	5550
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	5600
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	5650
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	5700
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	5750
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	5800
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	5850
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	5900
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	5950
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	6000
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	6050
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	6100
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	6150
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	6200
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	6250
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	6300
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	6350
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	6400
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	6450
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	6500
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	6550
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	6600
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	6650
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	6700
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	6750
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	6800
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	6850
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	6900
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	6950
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	7000
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	7050
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	7100
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	7150
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	7200
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	7250
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	7300
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	7350
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	7400
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	7450
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	7500
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	7550
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	7600
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	7650
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	7700
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	7750
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	7800
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	7850
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	7900
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	7950
GTGCTTTGCT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	AACTTTGCTT	8000

FIGURE 5B shows a sequence alignment of the DNA sequence of the invention with the DNA sequence of the reference sequence. The alignment is shown in a table format with the sequence of the invention in the first column and the sequence of the reference sequence in the second column. The alignment is shown for a portion of the sequence, from position 4250 to 8000. The alignment is shown in a table format with the sequence of the invention in the first column and the sequence of the reference sequence in the second column. The alignment is shown for a portion of the sequence, from position 4250 to 8000.

FIGURE 5D

TTTTACATAC	TTGGGATACG	TGAACTTTAT	TGATAGCTTC	TGAAAGCACT	TGCACATTTT	10660
ACATAATTTT	TACCTGCGTA	AGTGTATATG	TAAATCTTTT	GAATTTTAAA	TATACATGCG	10720
TTTGTCTAAG	TTTTTTTAAA	TGTTTATATG	TTTATCTAGA	TTGGATATCT	GGTATCTGCT	10780
GGATGACGTA	GAGGATACAT	TTTGGGACCT	TTGTGCGAGA	GCAATTTAAG	TTTGTGCTGC	10840
TGTGCACTGC	TTAGGGGAAA	AAATTTGCAAT	TGGGCTTTCA	ATTTTATTTT	TGTGAGGATG	10900
TTTGAGATTA	TATGTTGTAA	TTTGGTTGAT	AGATCTTTAT	TTATCTTTAT	TACAGCTTAT	10960
TTTGAGAGTT	TTGAGTATAT	TGATTTTAAG	TTTGACCTTA	GATGGGTGAA	TAGGAAGCTCA	11020
TAGAGCTTAA	TTTAACTTTA	TTTAAACGAT	ACATCTGATA	TTTGTGGATA	AACGGCTGTT	11080
TATGGGAGCT	TTTGTGTGTT	TGATGTTTGA	GCGGATGTTG	AGTTGTGCGA	CTCTAGTTGG	11140
GAACATTTT	TTAAATTTT	ACATTTGAT	TTGGGTAAAT	TTTAGAGCTAG	TTTGTGCTTAG	11200
AGCAATTTG	TTTAAATTTT	TTTATGTTTA	TGAGGGGTA	TTTAGAGGAT	AAATGCTCTT	11260
TTTGTGAGGA	AACTATTAAA	TTCTTCAAGG	AAAACTTAG	TTTATAGGAT	TATTTTATAA	11320
ATGTTTATTT	TTTGTAAGCA	TATTTGGGAC	ATACTTTAT	TTTGTGCTGT	TTCTTATGAG	11380
AACTATTAAA	AGTTATAGAA	AAATTTGTGT	TTTTGCTTT	ACTTAATGAT	AAATAATGAT	11440
TTAAATTTAG	TTTCTTTTAT	TTGGAGTATG	TGATTTTGTG	GTATTTTATG	AAACAACGAC	11500
GGTACAAATT	TTTAAATAGA	TAGGAGCTTT	GCAATGCGAG	AGGTTTGTCT	GTAGTCTCTT	11560
GGTTATATAT	ATTTTCTTCT	CTGATTTAGA	AGTTCTATAA	GGAAAGCTAA	GTATTTACAA	11620
TTGATAATGG	TTTAAATAGT	CAAGAGCGAG	AGATAGGATA	CTTTGCAAA	AATGACATTT	11680
ATTAAAAATG	TACTTTAAAA	TAGAGATATA	AAATTTGCTA	TTTAAATTAG	ATTAAGGTAA	11740
CACTTTATAT	TTTAAATACT	GAATGCTGTA	TGTAATGCTT	TTCTAGAGTT	GTATTTTATG	11800
AAAGGACAT	TTTGGAGCT	GAGGCGAGGA	AAATGGCTGA	ACGTTGGAGG	TGGAGGCTTC	11860
ACCTATTTT	TTTGTGCTTA	TGCGCTTCCA	AGTGGCGGTA	GACAGCGACA	TTTGTGCTTA	11920
ATTAATTTAA	AAATAGAAAG	ATATTTTATA	AAATAGTTCT	GAATATTTCT	TATATGAATA	11980
TATTTGAAAT	TTTGTGCTTA	TTTGTGCTTA	TGCACTTTAT	TACATTTTAA	ATGACATTTA	12040
AAATTTTCTT	TTTAAATTTT	TTTGTGCTTA	AGAAATTTTA	TACTAGATAT	TTTGTGCTTA	12100
CACTTTATAT	TTTGTGCTTA	TTTAAATTTT	AGCTTTTATA	AGAAATTTAT	TTTGTGCTTA	12160
TGCACTTTAT	AAATGATTTT	AGATGCTGTA	AAATGCTTAT	TTTGTGCTTA	AGCTTTTATA	12220
TTTGTGCTTA	TTTGTGCTTA	TTTAAATTTT	TAAATTTTAA	TTTGTGCTTA	TTTGTGCTTA	12280
TTTGTGCTTA	TTTGTGCTTA	TTTAAATTTT	AAAAAATA	ATTAAGCTTA	TTTAAGCTTA	12340
AGTTTCTTCT	TTTGTGCTTA	AAATTTAGAA	TTTAAATTTT	TTTAAATTTT	TACATTTTAT	12400
TTTAAATTTT	AAATTTTCTT	AAATTTAGAA	TTTAAATTTT	ATTTGCTTAA	TTTAAATTTT	12460
ATTAATTTTA	TTTAAATTTT	TTTGTGCTTA	TTTAAATTTT	GATTTGCTTA	TTTAAATTTT	12520
TTTGTGCTTA	AAATTTTCTT	AAATTTTCTT	TGAGAACTTA	TTTAAATTTT	TTTGTGCTTA	12580
AACTTTAGAA	AAATTTAGAA	TTTTTCTTAT	GATTTTCTAT	TTTAAATTTT	AACTTTTAAA	12640
TTTGTGCTTA	TTTGTGCTTA	TTTAAATTTT	TAAGATTTTA	AAATGATTTA	TAATGATTTA	12700
TTTGTGCTTA	TTTGTGCTTA	TAAGATTTTA	TTTAAATTTT	TTTAAATTTT	TTTAAATTTT	12760
TTTGTGCTTA	TTTGTGCTTA	TTTAAATTTT	TTTGTGCTTA	TTTAAATTTT	TTTAAATTTT	12820
TTTGTGCTTA	TTTGTGCTTA	TTTAAATTTT	TTTGTGCTTA	TTTAAATTTT	TTTAAATTTT	12880
TTTGTGCTTA	TTTGTGCTTA	TTTAAATTTT	TTTGTGCTTA	TTTAAATTTT	TTTAAATTTT	12940
TTTGTGCTTA	TTTGTGCTTA	TTTAAATTTT	TTTGTGCTTA	TTTAAATTTT	TTTAAATTTT	13000
TTTGTGCTTA	TTTGTGCTTA	TTTAAATTTT	TTTGTGCTTA	TTTAAATTTT	TTTAAATTTT	13060
TTTGTGCTTA	TTTGTGCTTA	TTTAAATTTT	TTTGTGCTTA	TTTAAATTTT	TTTAAATTTT	13120
TTTGTGCTTA	TTTGTGCTTA	TTTAAATTTT	TTTGTGCTTA	TTTAAATTTT	TTTAAATTTT	13180
TTTGTGCTTA	TTTGTGCTTA	TTTAAATTTT	TTTGTGCTTA	TTTAAATTTT	TTTAAATTTT	13240
TTTGTGCTTA	TTTGTGCTTA	TTTAAATTTT	TTTGTGCTTA	TTTAAATTTT	TTTAAATTTT	13300
TTTGTGCTTA	TTTGTGCTTA	TTTAAATTTT	TTTGTGCTTA	TTTAAATTTT	TTTAAATTTT	13360
TTTGTGCTTA	TTTGTGCTTA	TTTAAATTTT	TTTGTGCTTA	TTTAAATTTT	TTTAAATTTT	13420
TTTGTGCTTA	TTTGTGCTTA	TTTAAATTTT	TTTGTGCTTA	TTTAAATTTT	TTTAAATTTT	13480
TTTGTGCTTA	TTTGTGCTTA	TTTAAATTTT	TTTGTGCTTA	TTTAAATTTT	TTTAAATTTT	13540
TTTGTGCTTA	TTTGTGCTTA	TTTAAATTTT	TTTGTGCTTA	TTTAAATTTT	TTTAAATTTT	13600
TTTGTGCTTA	TTTGTGCTTA	TTTAAATTTT	TTTGTGCTTA	TTTAAATTTT	TTTAAATTTT	13660
TTTGTGCTTA	TTTGTGCTTA	TTTAAATTTT	TTTGTGCTTA	TTTAAATTTT	TTTAAATTTT	13720
TTTGTGCTTA						

FIGURE 5E

GAATTAAGCTT	TAAGCTTAAA	TGACCTTTCT	TAGCTTATAC	TGTTACTTTT	TGCGAGATTC	15860
AAGATAGAGC	AATGGAATTT	TTTCTCTCTA	TTTATGTTCT	TTTCTACTT	ATTGTGCTCT	15920
AAGATAGAGC	TTTAAATTTT	TATTTATGTT	TTGAAATGAG	TTGATAGGAT	AGAAGCTTTA	15980
TTTCAAGAGC	TTTGGGAAAG	CGAGGTGCGG	TTGATTTCTG	AGTCAAGGAG	TTTCAAGATG	16040
TTCTTGGGCA	CATATGCGAA	TTCTCTCTCT	ATATATATAT	TTTATATAT	TTTCTAGATG	16100
TTGATATGAT	TTCTCTCTAT	TTTCTCTCTG	TTGAGTTTCA	TTTGGGAGAT	CGATTTGAAT	16160
TGAGAAATTA	CGATCTCGAT	TAGCGATGAT	TACATCGATG	TTCTGAGAGG	TTGGGTGACG	16220
AGCAATACCG	TATCTCTGAT	GAATTAATGA	ATGATATGAT	GAATGAATGA	ATGCTCAATG	16280
TTCTTAAGCTA	TTTCTCTGAT	AGCAGCTTTT	TTCTATAGG	TTCTCTCTAG	TTCTCTCAAT	16340
CGACAACTTA	TTTAAATTAAG	TAAACACTTA	AGCAATTAAG	AGAGTAGGCG	TTCTTTCAAG	16400
TGAGAAATTA	TTTAACTTAA	TTTCTTTGAA	TTTCTTCTAG	TACTCTCTCT	TACTCTCTCT	16460
TTTCTCTCTT	TTTAAATGTA	TTTCTTTTAA	TTTGGATTTG	TAGTTTGAAA	TTGGAATATGA	16520
TTCTCTCTGA	AAGAAAGATC	TCTCTTTTAA	AGCAAAAGTT	TACTCTCTCT	TTCTATTTGA	16580
GAATCTTAGG	AACAGAGACA	TGTAAATTTA	TTCAAAATAG	AGCAAGTTCT	TATAGATTTA	16640
GTATCTTAGG	CGAAAGGACA	TTTAACTCTG	ATGATATGAG	TAAAGCTCTA	TTTCTTCTAT	16700
GTATCTTAGG	GTATTTTAA	TTCTTATGAA	CGAAAGGCTA	TTTGGGTAA	TTTCTAGTCT	16760
TAGAGAGCTA	GTATCTTAGG	TTAGGGATTA	TAAATACCTT	GAATTTATTA	TTTCTGCAAG	16820
AAATAGAAAT	ATGACTAGCA	AGGTGACCGG	TGACAGATGT	AGATTAAGCT	TAAAGAACTG	16880
TTCAAGGAAAG	GAAGAGCTAC	TAGATACCTA	TTCTATTTT	TTACTTTTAA	ATTAAAGAAA	16940
TGAGGTCTAG	TTTAAAGGATG	AGCTCTTAA	TAAATAAAGG	CTAAATATA	TTTACAGAGG	17000
AAGCTTTCTG	CTGAGGTTCT	AGCTTTTAA	CTACAGTTCT	TTTAAAGATA	AAATTTAATA	17060
TTCTCTCTAT	TTTAAATAGG	AAATAGAGCA	TAACTACTT	TTGAGAGATTA	TTGAGTTTAA	17120
TGAGGTCTAG	TAGCTTACAG	TTCTCTCTCT	GTATCTATAT	TTGAGTTTAA	TTGAGTTTAA	17180
TTCTCTCTGA	TTTCTCTCTG	ATGCTCTCTG	TATATATAGA	AAATTTTAA	ATTTATCTCT	17240
GTATCTTAGG	TACTTTTGA	TACTCTAGCA	TGACGATCTT	TTTCAAGGAG	GAGTTTAGAG	17300
CGATTTTAGG	CTAGATAGCA	TGAGCTCTCT	TTCTTAAAG	TACTCTTAA	TATATTTAGC	17360
CGATTTTAGG	GAATTAAGCT	TAGCTCTCTG	TTTCTCTAG	TTGAGTTCTG	TAGATTTCTT	17420
GAATTTTAGG	GAATTAAGCT	ATGCTCTCTG	ATGATTAAG	TTTCTCTAG	GAGCTTTCTT	17480
TAGTTTCTAG	GAATTAAGCT	ATGCTCTCTG	ATGATTAAG	TTTCTCTAG	TAGATTTCTT	17540
TTTCTCTCTT	AGCTTTCTTA	TTCTCTCTAG	TTGAGAGATG	AGCTTAAGCT	TTTAAATCTT	17600
TGAAATTTAG	ATGCTTTCTT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	17660
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	17720
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	17780
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	17840
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	17900
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	17960
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	18020
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	18080
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	18140
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	18200
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	18260
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	18320
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	18380
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	18440
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	18500
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	18560
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	18620
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	18680
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	18740
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	18800
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	18860
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	18920
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	18980
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	19040
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	19100
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	19160
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	19220
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	19280
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	19340
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	19400
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	19460
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	19520
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	19580
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	19640
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	19700
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	19760
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	19820
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	19880
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	19940
CGATTTCTG	TTTCTCTCT	GAATTTCTT	GAATTTCTT	TTTAAAGATG	TTTAAAGACT	20000

FIGURE 5F

[illegible]

FIGURE 5G

[illegible][illegible]

• • • •

[illegible]

• • • •

FIGURE 51

AGAGTGTGAT	TTATACAGCA	ATACAGCTAA	AGAGGAGGTT	AATAGATGCG	TAGTGCAGCT	11660
TTATGCTGCA	AAAGGCTACT	AGGCTGTAA	AAAGACACTT	AGAGTTTGA	JATTGCGAGG	11720
ATGAGGAGCT	AAATGCTGCG	TTAGAGCATG	GTGTTGCGAT	AGTGGGGTGA	CGAGTGTGAA	11780
TAGCGACAAA	TGCTGCTGTA	GTAAATGAGG	ATTGTTTCTT	GGTGGGAT	GTACAAATTA	11840
TAATATGAG	TTATATGAAA	GGTGTAAAAA	AGTAAACAGT	AAACATGTTA	TACCTGGAGC	11900
TTGTTTATTA	GGCGGAGCTG	ATGGTGAAG	ATGGAATATG	CGAGCTGTAG	GGTTATTTGA	11960
TTGAAATAT	TTTTTGTG	TATGTGAGAG	TTTAGGCTGT	AGATTCTTGG	ACATGTAAAA	12020
TTGAAATGCG	TAAGATATTT	GGGAGGACCA	TTTTATTTGG	TGGGGTTCAG	TATTTGAAAC	12080
ATTAGGCGCT	ATATATATAG	CATGTAAAA	GTAAAGACGA	AACATTTATG	TTTTGTGCGA	12140
AGCAATTAAT	TAGTATCTGA	GTAAATAAAT	TGTTATGAGC	GTATATATCA	GAATATGAAA	12200
ACAAAAATTA	AATGTTCTGA	AAAGAGAGCG	TATGTGAGCG	ATTGCTGTAT	ATATGTATAC	12260
AGATATATAT	AGAGACATCT	ATAAATTAAC	TATAAAAATG	ATAGCAATAT	CATTGATGTA	12320
GAAGAGAGCG	TGTTGTTTAA	ATTATTAATT	TATGCTGTGA	CAGGTGTTAT	TATGAGCTTA	12380
GTATGCTGCT	TTGATTTGCG	TTTTATGTGT	AGTTAAAAAA	ATAGTGTGTT	ATGCTTTTAC	12440
GTATGCTGCT	GGGTTTAAAA	AGATTTGTGT	ATGTTTGTAT	TAATCTGTTT	TTGAGTGTGT	12500
TGCTGCTGCT	TGTTTAAAG	ATTTGTTTAA	GAAGCATATT	TAAGATTTTA	TAGAAATATT	12560
AGATGTTTAA	TTGGGATGTA	TGATTTAAAA	TAGAGATATT	AGAGAGAGTG	TTATGATATA	12620
TTGCTGCTGA	ATATATAGAT	GGTTATAGTA	GTAAAAATTA	ATAGGTGATT	TTGTGCTGCT	12680
ATGATGCTAT	TGATGCTGTA	TTAGCTGCTA	GGAGATTTGG	GACTGTTTAT	AAGACAGAGT	12740
AGTTGCTGAT	ATAGTTGCTG	TTGTTAGCTG	TTGTATGCTG	TTAAAGAAAA	GAATAGGGAA	12800
TTGTTGCTAT	TGCAATATAT	TTATTTAAAG	TTGTTGCTGT	TTAGCATTAAT	TTTACCTTTT	12860
AAATATGATA	TGTTTAAAGT	GGAA	TTGAGGAGCT	GGTAAATGAT	TGAA	12920
GTGCTGCTAG	GGTTTAAAGT	GTGTTACTGT	TTGCTTAACT	GGTAATTAAT	ATGCAATGAT	12980
TGATTTGCTAG	TATGCTGCTAT	AGCTGCTGAT	TTAGGATGAT	TAAAGATGAT	TTGAGGATGAT	13040
TTGCTGCTGA	TTGCTGAGAG	TTGAGAGAGA	GAAGAGAGTA	TTGCTGCTAG	GAAGAGAGTA	13100
AGATGCTGAG	TTGCTGAGCT	ATGCTGCTGT	AGATGCTGTT	TATTTGAGCA	GAAGAGAGCT	13160
GGTGTGCTGA	TTGCTGATAT	CATGCTGCTG	GGAAATGCTG	TAGTGTGCTT	GAAGAGAGCT	13220
TTATATGCTG	TTGCTGATAT	GAAGTTGCTG	TTGCTGCTGA	ATTTGCTGTA	AGCTTACGAT	13280
AAATGCTGCT	TTGCTGATAT	TTGCTGAGTA	GGATGCTGCT	TTGCTGCTG	TTGCTGCTG	13340
TTGCTGCTAG	TATGCTGCTG	AGCTTACTGT	GGGAGCTGAT	ATGCTGTAGA	TATTTGCTG	13400
GGTGTGCTAT	TATGCTGCTG	TATGCTGCTG	GAATGCTGAT	TAGCTGATAT	TTGCTGCTG	13460
AGTAAAGCTT	TAAAGATGTA	AAATATGCTG	AGTGTGAGCA	AGCAATGCTA	AGATGCTGTA	13520
AGAAAAAGCA	AGGCTGAGAA	TAGAGAGAGA	AGCTTAAAG	TTGCTGCTTA	TGCACTGCTTA	13580
TATAGCTGCA	TATGCTGCTA	TGCTGAGAGAA	AAATATGCTG	TGCACTGCTT	TAAGCTGCTG	13640
GAAGCTGCTA	GAAGCTGCTA	TTTAAAGAT	TGCTGCTGAT	TGCACTGCTT	TAGTACTGCT	13700
TTGCTGCTG	TTTAAAGAGC	AGCTTACTGT	TTTAAAGAGT	GAAGCTGCTT	AGGCTGCTG	13760
TGCTGCTGCTA	TGCTGCTGAT	GGGAGAGCTT	GGGAGAGCTG	AGGAGAGAG	ATGATTTGAG	13820
GGGCTGCTAT	TGAGAGAGCT	TGCAAGAGCT	AGGAGAGCT	TTGCTGCTAT	AAAAATGCTA	13880
GGGATGCTAT	GGGCTGCTG	GGGCTGCTG	TAATGCTGAG	GGGCTGCTG	GGGCTGCTG	13940
GGGATTTGCTA	TGAAAGAGAG	AGGAGAGCT	TGCACTGCTG	TGAGATTTG	AGCTGCTGAT	14000
GGATTTGCTG	TGCTGCTGAT	AGGATTTGCT	TGCAAAAAAA	AAAAAAATTA	AGAGCTGCTG	14060
TGATGCTGCT	TGCTGCTGCT	AGATATGCT	TTTAAAGCTT	TTTAAAGCTT	TTTAAAGCTT	14120
TTTACTGCTG	ATGCTGCTG	GGCTGCTGTA	GGCTGCTGTA	GGCTGCTGTA	GGCTGCTGTA	14180
TATATTTGTA	TATTAATAG	AATTAAGCT	GGTAGCTTA	TTTAACTTAC	TGAAATAAGT	14240
GTGCTGCTG	AGAGTATGTT	TGCTTAAAG	TGCTTAAAG	GGCTGCTG	GAAGAGAGTA	14300
AAATATGCTA	TATATATG	TTTAAAGAT	AGATGATAT	AAATGGGAT	TTTAACTTAT	14360
GGGCTGCTG	TAGAGAGAG	TTGCTGCTA	CATGCTGCTA	ATGCTGTAT	ATAGCTAAT	14420
TGAGCTGCTA	TGCTATAGTAG	AAATGAAAA	AGTTACTTAA	AAAAATTTGT	ATTGGCTGCTA	14480
AAATTTGCTA	ATTAAATGTA	TTTCTTAAAG	TGAGAGATTA	TATGGGCTTT	CATGCTGCTA	14540
TTGATGCTAT	CTGCTGCTA	TTTAAAGAG	GTGCTGCTA	TTGCTGCTG	TTGCTGCTA	14600
TTATTTGCTT	GTATTTGCTT	GTGCTGCTA	GTGCTGCTAT	TATGCTGCTG	TTGCTGCTG	14660
TTGAGAGGCT	GTGCTGCTG	AACTGCTGAG	TTAAAGCTG	GTAGAGCTTA	TTTAACTGAG	14720
GGATACTAG	AACTTATGCT	AGGATTTG	AACTTATG	TGAA		14780
GGATTTGCTG	TGCTGCTGAT	TGAGAGAGCT	AACTTATG	TGAA		14840
TGAGCTGCTA	AACTTATGCT	TGAGAGAGCT	AACTTATG	TGAA		14900
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	14960
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	15020
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	15080
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	15140
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	15200
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	15260
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	15320
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	15380
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	15440
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	15500
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	15560
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	15620
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	15680
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	15740
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	15800
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	15860
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	15920
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	15980
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	16040
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	16100
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	16160
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	16220
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	16280
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	16340
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	16400
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	16460
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	16520
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	16580
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	16640
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	16700
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	16760
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	16820
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	16880
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	16940
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	17000
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	17060
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	17120
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	17180
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	17240
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	17300
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	17360
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	17420
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	17480
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	17540
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	17600
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	17660
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	17720
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	17780
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	17840
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	17900
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	17960
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	18020
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	18080
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	18140
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	18200
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	18260
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	18320
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	18380
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	18440
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	18500
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	18560
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	18620
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	18680
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	18740
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	18800
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	18860
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	18920
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	18980
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	19040
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	19100
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	19160
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	19220
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	19280
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	19340
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	19400
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	19460
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	19520
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	19580
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	19640
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	19700
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	19760
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	19820
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	19880
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	19940
GGATTTGCTG	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	TTTAAAGAT	20000

[illegible]

FIGURE 5J

ATTTTCTCTT	TTATGACGAT	TGGGATGCGA	TTATGCGGAG	GGGTGAAAGG	TGGGTGCGCG	17360
CTTCTATTTC	AGGCGAGCGG	CTTCTACTCA	GAGGTTGAGA	AGATTATTTG	CGATCTTTAT	17920
CGAAAGCTTT	CGGTCGAGTC	AGGTGAGAGG	GATGGCGTAG	CTCTGCGCGG	TTATGAGACT	17980
GATGAGCTTA	AGAGCGATTC	GACTATATTT	TTGACTGTTA	AGAGGATTTA	TGCTTTATTC	18040
TATTTATGAA	AGCTTACGTT	TGTGATTTTC	TTCTGCTTTT	TAGGAAAGCA	ATCTTTCTTC	18100
CGTTATGTC	TGAGAGGAAA	GTATATCTAG	AAATTTTTTT	GTCTTTCTTC	GTCTTTCTAG	18160
ACGAGCTTT	CTCTTTCTTC	TTAGGCTTGA	TTGCGCTGCG	GTGATCTTTC	CTGCTGCGCA	18220
CTGCTATCTT	CGAGCTTTCA	GTGATCTTTC	TGCTGCTGCT	TGCTGAGTAG	CTGGGACTAG	18280
AGGCGTCTTC	GACGATCTTC	AGCTACTTTT	TGTATTTTTT	GATAGAGACA	GGGTTTTTCA	18340
TGTTGCGCTG	CGAGCTCTTC	AACTCTGAG	TTCTGATGAT	CGGCGGAGCT	GAGCTTCTCA	18400
AACTGCTCTC	ATTAGAGGTC	TGAGCGATCG	GACCTGCGCA	ATACACTGAG	AAATTTTTAT	18460
TTCTCTCTTC	AGCTTAAGCT	TACGACTTTC	CGAGGATCTC	AAAGCTGCGC	CTGCTATTTT	18520
TTCTTAATTT	CTATCTGATC	ACTTGCAGAA	ACCTATTTTT	TCTGCGACTT	CACTTCTGCT	18580
AGCTTCTCTA	CTGCTAGTTC	TTCTCTAAAT	CGTCTGAGCT	CGTCTGCTAT	GGTTTTCTTC	18640
GAGTAGCTTC	TGTAATGATC	ACAGTCTATC	GTATGCTTTA	CTTGCAGCTC	GAAGGGACAG	18700
ACCAAGTTCT	CTGCGCGCTT	ACCTAGAGGG	ATTCTCTCTC	ACTTCTCTTC	AGAACTCTAG	18760
CTGCGCGGAG	CGAGCGCGCT	TGCTGCGCT	TGTAGAAATA	TTTTAAATTA	TTATCTCTTT	18820
TTTTTTTAA	AGAAATTAAT	AGGAGATAGC	TTAGAGGATT	TTCTCTCTTA	GATCTCTAAA	18880
TACAAACTTC	GGGTCTTATA	ACTGATTAAT	TCTGATTAAT	TTCTTTTGAC	TCTTAGGATA	18940
GAGGAGCTGC	GATACGATTA	CGCTGATCTC	CAAGGCTGCA	GTGAGATGAC	TTTTTACTAC	19000
CTTAAAGCTT	TTGCGATTTT	TGAAAGATTC	TTGAAAGATT	CTCTTAAATA	ATCTGCGAAG	19060
TGCTCTCTTC	AAAGCGAGCT	CAATACTCTA	CTTAAATGCA	TTCTTCTGCA	AACTAGCTGAC	19120
TTTAAAGATA	CGATTAATTT	TAGACTCTAG	TGCTGCTCTA	GTGAGGCTCT	AAAGAGTACA	19180
GAGTCTCTTC	AGCTGCTCTC	AGCGAGCTTC	GAAGGAGCTC	GAAGGAGCTC	GAGCTAGCTC	19240
AGTTGCGACT	TTCTCTCTTC	GAATTTGAGC	GATGCGCGCT	AGTTGCGGCT	CGAGCTCTTC	19300
TAGACTTAAT	GAATTAATTA	TGTAATGATT	ATGTAATGCT	GAATTTGCGA	ATAGGAGGCT	19360
AAATTAATTA	TGCTCTCTTC	TAGAGGCTTC	AGCTTAAGCT	CGTCTGATTC	GATATTAATA	19420
TGAGAGCTTC	TTGCTCTCTC	TGCTCTCTTC	TTGTAATGCT	AGGATTTCTC	GAGAGCTTAC	19480
TGAAAGCTTC	AGTTGCTCTC	AGGAATTTCA	GAGGAGCTTC	AGGAGGCTTC	TGCTGCTCTC	19540
ATTTCTTAAT	GAAGGAGGAT	AACTAGCTTC	AGGCTGCTTC	ATGAGGCTTC	AGCTGCTCTC	19600
AGTTGCTCTC	TTGAGGCTTC	AGGAGGCTTC	GAGGCTGCTC	TTGCTGCTTC	ATGAGGCTTC	19660
GTGAGGCTTC	GAGTCTCTTC	GAGGCTCTTC	GAGGCTGCTC	GTGAGGCTTC	GTGAGGCTTC	19720
GAGGAGCTTC	GAAGGAGGAT	ATGAGGCTTC	GTGAGGCTTC	GAGGAGCTTC	GAAGGAGGAT	19780
GTGAGGCTTC	GTGAGGCTTC	GAAGGAGGAT	GAAGGAGGAT	GTGAGGCTTC	GAGGAGCTTC	19840
ATTTGCTCTC	TGAGGAGGAT	TGAGGAGGAT	AGGCTGCTTC	GTGAGGCTTC	AGGAGGCTTC	19900
GAGGAGCTTC	GAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	19960
TAGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	20020
GAGGAGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	20080
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	20140
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	20200
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	20260
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	20320
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	20380
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	20440
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	20500
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	20560
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	20620
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	20680
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	20740
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	20800
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	20860
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	20920
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	20980
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	21040
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	21100
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	21160
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	21220
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	21280
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	21340
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	21400
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	21460
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	21520
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	21580
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	21640
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	21700
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	21760
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	21820
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	21880
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	21940
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	22000
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	22060
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	22120
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	22180
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	22240
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	22300
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	22360
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	22420
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	22480
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	22540
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	22600
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	22660
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	22720
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	22780
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	22840
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	22900
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	22960
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	23020
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	23080
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	23140
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	23200
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	23260
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	23320
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	23380
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	23440
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	23500
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	23560
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	23620
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	23680
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	23740
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	23800
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	23860
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	23920
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	23980
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	24040
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	24100
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	24160
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	24220
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	24280
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	24340
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	24400
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	24460
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	24520
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	24580
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	24640
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	24700
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	24760
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	24820
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	24880
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	24940
GTGAGGCTTC	GAAGGAGGAT	GTGAGGCTTC	GAAGGAGGAT	ATGAGGCTTC	GAAGGAGGAT	25000

FIGURE 5K

TTTGGGTTTGA	ATCGGAGCAC	TTTTGGGAGCG	CGAGGCGAGCG	AGATGCGTTG	AGGTCGGGAG	42060
TTTGAAGAGCA	CGCTGGCGAA	GATGATGAAA	CTCGGTTTTCT	ACTAAAAAGTA	CAAAAAATTAG	42120
TTTGGGCTTGG	TTTTGGGCTGC	TTTGAATCTCG	AGCTGCTTAGG	GAGGCTTGAGG	CAGGAGGATT	42180
CGTTTGAAAGC	AGGAGGTTGGA	GGTTGCGAGTC	AGTCAAGATT	CTCGGCGAGT	ACTCTGAGCT	42240
GGGAGACAGA	CGGAGACTTT	GTCTTGAAAA	AAAAAGAAAA	AAAAAGAGA	TAAAAAGGTA	42300
TAATGCTAGA	GGGCACTTAC	CGCGAATGGA	GGTTGCGAGCT	TGGGAGTTTGC	TCTGGGTTAG	42360
TGAGTGAAGT	AGGCGTTGAG	GAACTGAAAG	ACGTAGGAGT	GTGCGCTGCT	GTAGAGCTTA	42420
TAAAGCTGCT	CGACTGAGGC	GACGCTGAG	CTGTGCTATAC	GAGTCTGAGT	ACTGTATTAAC	42480
GTAGCTTGCG	ATGTAGCTTT	GAACTGAAA	CAAAAGTTAA	AAAAATTATG	TTCTTTTGGC	42540
TATATTAAT	TAAAGCTTGG	TACTGTTTCT	GATTTTCTTT	TATGAGTTAA	AACTTTTATA	42600
CTGCTGTTGA	ATTAAGCTGTC	GGTTAAAAAC	CAAACTATTT	GTACAGCTAT	ACAAATATAT	42660
TTTCTTTTGA	TGCTTTCTTT	TTAAGATTTT	TTCTGTTTTT	GATTTTCTTT	AACTTTCTTT	42720
TACTTTTATG	ATTTTCTTTT	TTAAAAAGCA	AGACAAAAAC	CGACAGCTGA	CGCTTAGGCT	42780
ACATGGGCTTG	AGGATGATTA	GTCTGACTAT	TTTGCAGCTT	CAGATCTTGT	CGGACGAGGT	42840
CTTGAGGGGCG	AGTCATATTC	ATGGGGCTGT	CGTCTGCTGT	GATACAAAGT	CGCTGCTGTG	42900
GACAGCTTGA	GAAGGGCTTG	CGTTTTTTAC	AGTGAAGCTG	TAAAAATATA	TAAATATGAT	42960
AGTATGAGTA	ACATCAAAAC	ATATTAACAT	AGTCACTTCT	TATCATTTT	AACTATTTATA	43020
TACTGTACAT	AACTGTGAGT	GTGAGAGAGT	TACACAGCTT	CGAGCGAGGT	GAGTTTGTGT	43080
ACACTATTAC	CGCGACAAAC	ACATGGGTTA	TGCTTTGACT	TGTGATGTTA	CGATGGGCTG	43140
ATGTCCTTGA	GTGCTAGGAA	GTCTTGCTCT	TAGCTATATC	GTAACTGGATA	CTTTCTGCTG	43200
TTTGGCTGCG	GTCTTGCTAC	CGAACTGAGT	TATGCTGTGC	ATGCACTGTA	ATTAGATAGT	43260
CTTGAGAAAG	CTTGCGGAGC	TTTGCTAATAG	TAAATGCTGG	TGGCAAGAT	GATGATGATG	43320
ATGATGTCGA	TTGAGAGACT	AGATGTTAAA	ATTTTATGCT	GTCTTAAAGC	TAGCTTTTAA	43380
ATAGCATTTA	TTGATATGTT	TGTTTCTGCT	TATAGCTGCT	TACCAATCTT	TGCAATTTAT	43440
AGCTATTTAA	GAAGAGATTT	TGCGAAAAAC	ATAGCTTTGA	CGGCGATAGA	ACTTGAAGAT	43500
TGAAAGCGCT	TGTAAGCTTT	TGTTGCTTGG	TGTCGCGAGC	TGCGCGAGTC	CTGCTGCTTT	43560
GACATTTAAG	AACTGCTGGG	GATATGTTGG	TGAACTGCTG	AAATCGAGAG	TGTAAGCTTA	43620
TTTCACTTGG	TGCGAGCGGA	TAATGCTAAG	AAAAATTTTA	TAGATATCTA	TGACGCGAGA	43680
TATTTATAGG	AGCTGCTGAG	TAATGCTAAG	TTAAAGATTT	GAGATCTGCT	TAACTTTGCT	43740
TGAAAGCTGT	GTAGCGGAGTA	GTATATGCGA	TTTATAGATT	AGAAATGAAA	GTGGAATGTA	43800
TTAGAGCTTT	CGTTTCTGAT	CGGAGCTTTG	TTTTGCGAGC	AAAGAGCTTG	AGCTTTAGAG	43860
AGCTATTAAG	GTGTTGCTAT	AGGATCTGTT	AGATTAATTA	GTGAGAGAGT	TGCTCTGCTT	43920
AGTCTGTTTA	AGATTTGACT	CGCAATTTAG	TTAAAGAGAA	GAGATATGCA	ATTGTTAATG	43980
TTGATATATC	AAATGCTGAT	ATAGATTTTA	AGCTGATGAG	TGAGAGAGAT	TATTTGTTAT	44040
AGCACTATTT	TGAGATGCTT	AGTATTCTAT	TGTTGTTTAA	TGCTGCTTTC	TTTAAAGCTA	44100
AGAAATTTTA	AAATTTAGAG	GAATATTTTA	CTAAAAAAAT	GTAGAGATTT	ACTTATTTAA	44160
AGCTGCTGCT	TGCTTTTACT	GACTGCTGCT	AGGCTGTGCA	TGAAATATTA	AAAAATATAT	44220
TTTCTGCTGG	TATTTTCTAA	TGTAGAGAGT	AGAGCTTTTA	TAGAGCTGCT	ATGAGCTGCT	44280
TTGCTGCTGG	TGCTGCTGCT	GAGAGAGGCT	TTGCTGCTGT	TGCTGCTGCT	GTGCTGCTGCT	44340
TGCGAGCTGA	TAGCTTATGT	TAGGCTTTAA	GTCTTGCGCT	CGAGCGAGCT	TTTGCAGCTTA	44400
GGTTGCTTAA	TGCTGCTGAG	CGAGCTTTGG	TAAATGCTTTA	GTATGCTGCT	AGAGATGAGT	44460
TTGAGAGTGA	TTGCGCGAGC	TGCTGCTGTA	CGGCTAGGCT	TAGAGGATTT	TGCGAGCTTA	44520
GGCTGCTTAA	GTCTGCTGAT	TAGAGCTTGT	AGTGTGTTAG	CGAGCGGCTT	TTTATTTTCT	44580
TTTCAATTTA	GAGTCAAGAT	GATTAAGCTT	TGAATGCTGG	AGCTGAGATT	GTGCTGCTGCT	44640
ATCTTATGCT	TTCTTTTAGA	AAATATCTTC	CGTTCTTTTA	CTGAGAGAAA	GATTTGTTGA	44700
TTATTAATTA	TATTAAGAGT	AAATGAGGTA	TAGATATAT	AAATAGAGAT	GAGATATGTA	44760
ATCTTATGAT	TGTAATTTAG	GTGAGGATTC	CGAGATTTTC	TGAGAGTTTT	AAAGAGAGAC	44820
TTTGATTTAA	ATCTAAATCT	TGAGCAATTT	TGAATTTAGA	AGTGTGCTTT	ATTTTATTTT	44880
GAAGAGATAT	TATGTTAGAA	ATAGAGATAT	TGAGAAATTA	CGAAAGCTAG	GTATATATAT	44940
GATAGCTTGA	TTGTTTAAAT	AGATTTGCTG	TGGAAAGCTG	TTGCTATCTA	AAAGAGCTTT	45000
GTAGAGCTGT	ATGCTGCTTC	TTTATGCTGCT	TTTATAGATTT	TTTAAAGCTT	GTGCTAGGAT	45060
TGCGATGCT	AAATATGCTA	AAATATTAAT	TAAAGGCTTT	AACTGCTTAA	GTAGAGCTTA	45120
TTTGTATTTA	TTTATTTGCT	CGATTTTCTA	TGCTGCTGTA	AGCTGAGATT	TAGCTATGCT	45180
TGCG						

	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100	2101	2102	2103	2104	2105	2106	2107	2108	2109	2110	2111	2112	2113	2114	2115	2116	2117	2118	2119	2120	2121	2122	2123	2124	2125	2126	2127	2128	2129	2130	2131	2132	2133	2134	2135	2136	2137	2138	2139	2140	2141	2142	2143	2144	2145	2146	2147	2148	2149	2150	2151	2152	2153	2154	2155	2156	2157	2158	2159	2160	2161	2162	2163	2164	2165	2166	2167	2168	2169	2170	2171	2172	2173	2174	2175	2176	2177	2178	2179	2180	2181	2182	2183	2184	2185	2186	2187	2188	2189	2190	2191	2192	2193	2194	2195	2196	2197	2198	2199	2200	2201	2202	2203	2204	2205	2206	2207	2208	2209	2210	2211	2212	2213	2214	2215	2216	2217	2218	2219	2220	2221	2222	2223	2224	2225	2226	2227	2228	2229	2230	2231	2232	2233	2234	2235	2236	2237	2238	2239	2240	2241	2242	2243	2244	2245	2246	2247	2248	2249	2250	2251	2252	2253	2254	2255	2256	2257	2258	2259	2260	2261	2262	2263	2264	2265	2266	2267	2268	2269	2270	2271	2272	2273	2274	2275	2276	2277	2278	2279	2280	2281	2282	2283	2284	2285	2286	2287	2288	2289	2290	2291	2292	2293	2294	2295	2296	2297	2298	2299	2300	2301	2302	2303	2304	2305	2306	2307	2308	2309	2310	2311	2312	2313	2314	2315	2316	2317	2318	2319	2320	2321	2322	2323	2324	2325	2326	2327	2328	2329	2330	2331	2332	2333	2334	2335	2336	2337	2338	2339	2340	2341	2342	2343	2344	2345	2346	2347	2348	2349	2350	2351	2352	2353	2354	2355	2356	2357	2358	2359	2360	2361	2362	2363	2364	2365	2366	2367	2368	2369	2370	2371	2372	2373	2374	2375	2376	2377	2378	2379	2380	2381	2382	2383	2384	2385	2386	2387	2388	2389	2390	2391	2392	2393	2394	2395	2396	2397	2398	2399	2400	2401	2402	2403	2404	2405	2406	2407	2408	2409	2410	2411	2412	2413	2414	2415	2416	2417	2418	2419	2420	2421	2422	2423	2424	2425	2426	2427	2428	2429	2430	2431	2432	2433	2434	2435	2436	2437	2438	2439	2440	2441	2442	2443	2444	2445	2446	2447	2448	2449	2450	2451	2452	2453	2454	2455	2456	2457	2458	2459	2460	2461	2462	2463	2464	2465	2466	2467	2468	2469	2470	2471	2472	2473	2474	2475	2476	2477	2478	2479	2480	2481	2482	2483	2484	2485	2486	2487	2488	2489	2490	2491	2492	2493	2494	2495	2496	2497	2498	2499	2500	2501	2502	2503	2504	2505	2506	2507	2508	2509	2510	2511	2512	2513	2514	2515	2516	2517	2518	2519	2520	2521	2522	2523	2524	2525	2526	2527	2528	2529	2530	2531	2532	2533	2534	2535	2536	2537	2538	2539	2540	2541	2542	2543	2544	2545	2546	2547	2548	2549	2550	2551	2552	2553	2554	2555	2556	2557	2558	2559	2560	2561	2562	2563	2564	2565	2566	2567	2568	2569	2570	2571	2572	2573	2574	2575	2576	2577	2578	2579	2580	2581	2582	2583	2584	2585	2586	2587	2588	2589	2590	2591	2592	2593	2594	2595	2596	2597	2598	2599	2600	2601	2602	2603	2604	2605	2606	2607	2608	2609	2610	2611	2612	2613	2614	2615	2616	2617	2618	2619	2620	2621	2622	2623	2624	2625	2626	2627	2628	2629	2630	2631	2632	2633	2634	2635	2636	2637	2638	2639	2640	2641	2642	2643	2644	2645	2646	2647	2648	2649	2650	2651	2652	2653	2654	2655	2656	2657	2658	2659	2660	2661	2662	2663	2664	2665	2666	2667	2668	2669	2670	2671	2672	2673	2674	2675	2676	2677	2678	2679	2680	2681	2682	2683	2684	2685	2686	2687	2688	2689	2690	2691	2692	2693	2694	2695	2696	2697	2698	2699	2700	2701	2702	2703	2704	2705	2706	2707	2708	2709	2710	2711	2712	2713	2714	2715	2716	2717	2718	2719	2720	2721	2722	2723	2724	2725	2726	2727	2728	2729	2730	2731	2732	2733	2734	2735	2736	2737	2738	2739	2740	2741	2742	2743	2744	2745	2746	2747	2748	2749	2750	2751	2752	2753	2754	2755	2756	2757	2758	2759	2760	2761	2762	2763	2764	2765	2766	2767	2768	2769	2770	2771	2772	2773	2774	2775	2776	2777	2778	2779	2780	2781	2782	2783	2784	2785	2786	2787	2788	2789	2790	2791	2792	2793	2794	2795	2796	2797	2798	2799	2800	2801	2802	2803	2804	2805	2806	2807	2808	2809	2810	2811	2812	2813	2814	2815	2816	2817	2818	2819	2820	2821	2822	2823	2824	2825	2826	2827	2828	2829	2830	2831	2832	2833	2834	2835	2836	2837	2838	2839	2840	2841	2842	2843	2844	2845	2846	2847	2848	2849	2850	2851	2852	2853	2854	2855	2856	2857	2858	2859	2860	2861	2862	2863	2864	2865	2866	2867	2868	2869	2870	2871	2872	2873	2874	2875	2876	2877	2878	2879	2880	2881	2882	2883	2884	2885	2886	2887	2888	2889	2890	2891	2892	2893	2894	2895	2896	2897	2898	2899	2900	2901	2902	2903	2904	2905	2906	2907	2908	2909	2910	2911	2912	2913	2914	2915	2916	2917	2918	2919	2920	2921	2922	2923	2924	2925	2926	2927	2928	2929	2930	2931	2932	2933	2934	2935	2936	2937	2938	2939	2940	2941	2942	2943	2944	2945	2946	2947	2948	2949	2950	2951	2952	2953	2954	2955	2956	2957	2958	2959	2960	2961	2962	2963	2964	2965	2966	2967	2968	2969	2970	2971	2972	2973	2974	2975	2976	2977	2978	2979	2980	2981	2982	2983	2984	2985	2986	2987	2988	2989	2990	2991	2992	2993	2994	2995	2996	2997	2998	2999	3000	3001	3002	3003	3004	3005	3006	3007	3008	3009	3010	3011	3012	3013	3014	3015	3016	3017	3018	3019	3020	3021	3022	3023	3024	3025	3026	3027	3028	3029	3030	3031	3032	3033	3034	3035	3036	3037	3038	3039	3040	3041	3042	3043	3044	3045	3046	3047	3048	3049	3050	3051	3052	3053	3054	3055	3056	3057	3058	3059	3060	3061	3062	3063	3064	3065	3066	3067	3068	3069	3070	3071	3072	3073	3074	3075	3076	3077	3078	3079	3080	3081	3082	3083	3084	3085	3086	3087	3088	3089	3090	3091	3092	3093	3094	3095	3096	3097	3098	3099	3100	3101	3102	3103	3104	3105	3106	3107	3108	3109	3110	3111	3112	3113	3114	3115	3116	3117	3118	3119	3120	3121	3122	3123	3124	3125	3126	3127	3128	3129	3130	3131	3132	3133	3134	3135	3136	3137	3138	3139	3140	3141	3142	3143	3144	3145	3146	3147	3148	3149	3150	3151	3152	3153	3154	3155	3156	3157	3158	3159	3160	3161	3162	3163	3164	3165	3166	3167	3168	3169	3170	3171	3172	3173	3174	3175	3176	3177	3178	3179	3180	3181	3182	3183	3184	3185	3186	3187	3188	3189	3190	3191	3192	3193	3194	3195	3196	3197	3198	3199	3200	3201	3202	3203	3204	3205	3206	3207	3208	3209	3210	3211	3212	3213	3214	3215	3216	3217	3218	3219	3220	3221	3222	3223	3224	3225	3226	3227	3228	3229	3230	3231	3232	3233	3234	3235	3236	3237	3238	3239	3240	3241	3242	3243	3244	3245	3246	3247	3248	3249	3250	3251	3252	3253	3254	3255	3256	3257	3258	3259	3260	3261	3262	3263	3264	3265	3266	3267	3268	3269	3270	3271	3272	3273	3274	3275	3276	3277	3278	3279	3280	3281	3282	3283	3284	3285	3286	3287	3288	3289	3290	3291	3292	3293	3294	3295	3296	3297	3298	3299	3300	3301	3302	3303	3304	3305	3306	3307	3308	3309	3310	3311	3312	3313	3314	3315	3316	3317	3318	3319	3320	3321	3322	3323	3324	3325	3326	3327	3328	3329	3330	3331	3332	3333	3334	3335	3336	3337	3338	3339	3340	3341	3342	3343	3344	3345	3346	3347	3348	3349	3350	3351	3352	3353	3354
--	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

FIGURE 5L

AAATTAATTTG	GGATACATAC	ATACACACAC	ATACATACAC	ATTAATAAAT	GATGAATATG	46250
TAAATTTGGG	GGATATTTCA	GGATATTTCA	ATGATATTAAT	GACCAAGAGA	CAATATACAT	46320
TTTATGTTGG	TTGAGAGATG	GGATCATACG	GTTCCTGATC	TTACTATGTA	TAGAGGTTTC	46380
CTTTGCTTTT	GTAAATTTT	GAATCTTTTA	TTAGTTTAACT	CACTTACGTA	TTTCTTTTAA	46440
GGCTGTAAAC	TTTCTTTTGG	TTTCTTTTGG	TTTCTTTTGG	TTTCTTTTGG	GGTCTTTTAA	46500
TGCAATATTT	TGAGGATGCT	TGAGGATGAC	AGCTAATTTT	TGGAAGAAAG	GAATATATAA	46560
AGCTCAAAAC	TAAATATTTA	AAGAACTGCT	TTTCTAGGTTG	TGATTTGAAA	AATATCTTGA	46620
TTTCTGATCT	AGCTGATTTG	NAATGATGCT	CAATATTTTG	ATAGCTTTTG	TAAATTAAGT	46680
GAAGATGAGT	GAATGATGCT	TAACTGCTCT	GGCTTTACCG	CACTCTGGAG	CACTTCCAGG	46740
GAATGATGCT	TTGGAAGGAA	GTCTCTGAAAC	ATTAATTTTG	GGTCTGCTAT	AGGAGGTTTC	46800
TACATTAATTT	TAGAGGTTTC	GAAGAAATTC	ACACATGCTG	NAATGATGCT	GTCTGATTTG	46860
AAAAATGCTA	AACTTCTGCT	ATCTCTGAAA	TAGAAATAT	GTCTCTGAGG	ATCTCTGCTA	46920
GTAAATGCTG	TTTAAATGCT	ATCTGATGCT	GTCTCTGCTA	CAATATTAAG	CACTTAAAGA	46980
ATCTCTGCTA	TTAAATGCTG	ATCTCTGCTG	AGCAATTAAGT	TCTATTTTAA	TAAATATTTT	47040
TTATATCTGA	TTTAAAGGAA	AAGGCTACAC	AGTGATGAGA	ATTAATTTAT	TTAAAGATTA	47100
AACTTAATTT	AACTGATAT	TGGAATTTGA	AGCAATTTT	GTCTCTGAGG	AACTCTGCTA	47160
ATCTCTGCTG	TGATTTCTGA	GGCACTCTAC	AAAAAGACCT	TTGAAACACA	ATTTATGCTG	47220
TGATTAATTT	AGCTGATGCT	GTCTCTGCTG	TGGAAGGAGG	GGCTCTGAGT	GATATCTGAG	47280
TGGAATTTTT	AGCTGATGCT	CACTCTGCTA	GGCTATAGGG	ATTTGAAAGA	GGAAATTTT	47340
GTCTCTGCTG	GTCTCTGCTG	GTCTCTGCTA	GAATATTTTA	GTATTTCTAG	GTCTATGCTT	47400
TTTAAATTTA	TTTAAATTTA	TTTAAATTTA	AACTGATTTT	TTTAAATTTT	TACAGATATG	47460
AGTAAATTTA	GGTAAATTTG	AACTGATTTT	TTTAAATTTT	TTTAAATTTT	TTTAAATTTT	47520
TTTAAATTTT	AGATGCTTTA	AGATGCTTTT	GAGAGGCTTA	TTTAAATTTT	ATCTGATTTT	47580
GAAGAACTGA	GATTTTTTCT	GGCTCTGCTA	GAGATTTCTT	TAGTTCTTAA	GAGAACTGAG	47640
GAAGAACTGA	GGCTCTGCTA	GGCTCTGCTA	ATTAATTTTA	GAATTTCTGA	GAATTTCTGA	47700
TTTCTCTGAG	TGGAAGGCTG	TAGTAATTTA	ATCTCTGCTA	TTTCTCTGAG	TTTCTCTGAG	47760
TTTCTCTGAG	TAGTAATTTA	ATCTCTGCTA	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	47820
GAAGAACTGA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	47880
TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	47940
TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	48000
GAAGAACTGA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	48060
TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	48120
GAAGAACTGA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	48180
TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	48240
GAAGAACTGA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	48300
TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	48360
GAAGAACTGA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	48420
TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	48480
GAAGAACTGA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	48540
TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	48600
GAAGAACTGA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	48660
TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	48720
GAAGAACTGA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	48780
TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	48840
GAAGAACTGA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	48900
TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	48960
GAAGAACTGA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	49020
TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	49080
GAAGAACTGA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	49140
TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	49200
GAAGAACTGA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	49260
TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	49320
GAAGAACTGA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	49380
TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	49440
GAAGAACTGA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	49500
TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	49560
GAAGAACTGA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	49620
TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	49680
GAAGAACTGA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	49740
TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	49800
GAAGAACTGA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	49860
TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	49920
GAAGAACTGA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	GGCTCTGCTA	49980
TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	TTTCTCTGAG	50040

FIGURE 5M

ATGCTATATG	CAATTAATG	TCTGATATG	TTCTTTTTC	AGTGTATGCG	TCTGGGCTAT	50460
CTTGGGGTTC	TCTGTTATG	AGACAGCTTC	CTTCTGGGTC	CTTCTGCTTC	CTTCTACAGAT	50520
GTCACTACTG	TGCAAGTTC	AATTCCTGTC	GGTGGGTTGC	CTCTACATTC	TTACAGTTTC	50580
AGTCTGGTGA	GGATACCTTC	AAACTCTGTC	TTATTTGTAA	TCTGGACATG	GGATTTTGGG	50640
TTTCTACTAT	TACTTTCTTC	TTAATTTCTG	AATTTGAGAA	GATTAAGAGG	TTCTTTCTTC	50700
CACTCTCTTC	CACTCTCTTC	CACTACCTTC	TTCTCTATG	TCATTTTTTC	CTTTTATTTT	50760
TTTCTGACTG	TATTAAGAGG	AATGTATGAC	ATTTCTGCTC	TCACCTCTGA	CTTTGATTTT	50820
AAATTAAGAT	ACACATATTC	TTATACAAAT	TTTCTTCTAG	AAGATTTTAT	TACAGATGCT	50880
GATTTACAGG	TAAAAATTAC	TTATGAAATC	AGTTTCTGTC	ACCAATTTAT	CACTCTCTCT	50940
AACTTAATAT	ATGGATTTGAT	CTTCTTTGAT	AATGAAATTA	AATGTGATAG	TAACTTACAT	51000
ATTTCTGATG	TTCTTTACAT	CTCTATCTAC	ACATATATTC	ACCAATGATG	TTCTCTCTTC	51060
CTATTTCTTA	CTTAATTTGTA	ATAGCTTGGT	AACAGAGCTG	GGAGTATTTA	AAAGATTTAA	51120
AGATCTCTTC	AAATTTCTCT	TTCTTGGGAT	TTTATGTTTC	TACTGATGAA	GGAAATAGAC	51180
AATGGAAGCT	CTTCTTTCTA	TTAGGTAATC	TAGATATGAT	ACTGAGAGAT	TGAAATATCT	51240
AATTTCTGAC	CTCAAAAGAG	ACACTTATGTC	TAACTAAGCA	TTTCTCTCTC	TTTCTGATTC	51300
AAATTAAGAT	ATTATTTATTA	TTCTATTTAT	AATGCTGATG	ATCTATATTA	TAGAAATATA	51360
GAAGATAGAG	CTTGAATGTA	TATTTCTGAC	CAATTAATGA	CTTTGAGGAA	GAATTCAGGA	51420
TCTCTCTCTG	TATTTCTACAT	TTATTTCTAT	TTAACTCTGA	AGAAATTTGAT	GATGTTTACCT	51480
ACTATTTATTC	CAATTTCTCT	CTGAGAGAGCT	TGAAGCTTTA	GTAAAGCTGCA	TAAATAAGGTC	51540
ATACATTTAG	CAAGTTCTCT	AAATTAAGAT	CAAGCTCTCT	TCTGCTCTCT	TTTAAAGGCT	51600
CTCTATTTTC	ATGCTATTTAG	CTCTACAGAT	GAATTAAGGCT	TTCTCTCTCT	CTCTATTTCT	51660
TCTTCTAGAT	TACTCTCTCT	TTCTCTAGAT	GGGAATTAAT	TTTCTGATTC	TATCTATCTC	51720
AGAAAGCTTC	AATATCTGCA	ATTTACAGAG	AACTTAAGCA	AATTTACAGG	AAAAAAGAAA	51780
GAATTTTAAA	AATTTCTGCA	AACTTTGATC	TACAGAGATC	TGAAAGAGAG	AGATTTAGCT	51840
AGCTTAAGAA	CTATGATGTA	AACTTTGATC	ATGCTGATTC	ATGCAAGATG	TGCAAGATTC	51900
CTTTCTCTCT	TTCTCTCTCT	TATCTCTGAT	ATTTATGATC	TCTAACTGAT	AGTACCTGAT	51960
AGCTGATATA	AGCTGATATA	CTTCTCTCTA	TTTCTCTCTA	CTTCTCTCTA	ATGCTCTGAT	52020
TACCAAGAAA	CTTCTCTCTA	CTTCTCTCTA	AACTCTCTAG	AGCTAACTTC	CTTCTAAACT	52080
CTCTCTCTCT	TACTCTCTCT	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	52140
TAATTTATTA	ATTTCTCACT	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	52200
ATTTCTCACT	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	52260
AATTTCTCTA	TACTCTCTCT	TATTTCTGAT	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	52320
AGACAGAGCT	TCTTCTCTCT	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	52380
TCTTCTCTCT	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	52440
CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	52500
CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	52560
CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	52620
CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	52680
CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	52740
CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	52800
CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	52860
CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	52920
CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	52980
CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	53040
CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	53100
CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	53160
CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	53220
CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	53280
CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	53340
CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	53400
CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	53460
CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	53520
CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	53580
CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	53640
CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	53700
CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	53760
CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	53820
CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	53880
CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	53940
CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	CTTCTCTCTA	54000

FIGURE 5P

[illegible]

FIGURE 5Q

AGGCTGCTGAG	TTTATGAGAG	GATGGGCTTG	GTATGCTGAG	GGCAGGCTGG	TGTGGAACTG	672960
GTGACTGCTG	GTGATGCTGG	TGGGTGAGCG	TGGGAAAGTG	TTAGGATGAG	AGGGGCTGAG	672965
AGGGTGGGGA	GTGATGAGG	AGGTATTGAG	TGAACTGAGG	TGGGGTGTGT	GTGATGAGGG	672970
AATGAGCGAG	GAGGACAGGG	GTGGGCTGGG	TTGAGGATCT	AGGGGCTGAG	TGTGCTGTGG	672975
ATGCTGCTAG	GAGGCTGGAG	AGTTGCTTGT	TATGCTGTG	TTGCTGCTGG	TGTGCTGTGG	672980
AGAGCTGAGG	TGTAGCGAGT	AGATGCTTGG	ATGCTGCTGG	TTGCTGAAAT	TGGGCTGTAT	672985
AGGCAAAATG	TGAGCAATAT	TTGTATGATG	TTGCTTGACA	TTGCTGCTGG	AGGTATGATG	672990
TAGATTTGTT	TTTAATGCTG	TGATGCTTGA	GGCTGCTGAT	GAGATGTGAT	GTGCTGAGAG	672995
AGGCTGCTGT	TGATGAGCAT	AGCAAAAAGG	GTTCAGAGCG	TGCTACTGTT	TATGCTGAGG	673000
TATGCTGTGT	GTGCTGTGAG	TGCTGAGGTT	GTGCTATGAT	GTGCTGCTGT	TGCTGCTGCT	673005
GTGATGCTGT	AGGCTGAGGG	TAGATGCTAT	GATATGCTGT	TTATGAGAT	GTATATATG	673010
TGGCTGAGTG	AATGAGAGCA	TAGATATTTT	AGTTGAAATG	AATTTAAATG	AATTTAGAT	673015
GAATAGAGCA	GATGCTGTGG	TGAGCGCTGT	AATGCTGAGC	TTTGTGGAGG	GTGAGGGGCG	673020
TGGATGTGCT	TAGGTGTAGG	GTGCTGGAGG	AGGCTGAGCA	AGCTGTGAA	AGGCTGCTG	673025
TAGTAAAAAT	AGAAAAATTA	GGGGGCTGTG	GTGGGAGGCA	GTGCTGAAAT	GAGCTACTGT	673030
GGAGGGCTGG	GCGAGGAGAT	TACTGTGAGG	TGGAGGGCTG	AGGTTGCAAT	GAGCTAGCAT	673035
TGAGGCTGCT	GACTGAGAGT	TGGGCAAAAG	AGTGACACTG	TTGCTGAAAT	AAAGAGAAAT	673040
AGGCTATGCT	GGCTGTGGGG	TATGTATATG	GAGAGGAGAG	TTGCTGTGTT	GGATGTAGAA	673045
TGGATGTGTA	TGAGGGTGGG	TATGCTGAGT	TATGTAGAGG	TAATATGCTT	GAATGTAGAA	673050
TAGCTGTGAT	GTGCTGAATG	TTGCTGTGTT	TTGCTGAAAT	TATGCTGTGG	GTGATGCTGG	673055
TATATGAGAG	TTTAAAGAGG	TTTGTAGAGG	TATATAGGCT	ATGAGGGGAG	AGTCTGCTAG	673060
GAGAGGCTAT	AGGCTGTGAG	AAGAGCAATG	GAGTGTGCTA	TGAGGAGGAG	TAGCACTGCTG	673065
GTGCTGCTGT	TTGCTGTGGG	AGGCTTAGCA	GAGGCTGTGT	GTGGGCTGGG	AAGAGAGGAG	673070
TGAGATATGT	GTGCTGTGCA	AGGAGGAGAG	AGGCTGCTGT	AGAGGCTGCT	GTGATGCTGG	673075
GTGCTGCTGT	GAGCTGTGAG	GTGCTGTGAG	TATGTAGAAAT	GAATATGCTT	TTGCTGTGAG	673080
TTGCTGTGAG	TTGCTGTGAG	AGGCTGAGAG	GAGCTGAGCA	TATGCTGCTG	TATGCTGCTG	673085
TTGCTGTGAG	TTGCTGTGAG	AGGCTGAGAG	TTGCTGAGCT	TGAGGAGGAG	TATGCTGCTG	673090
TTGCTGTGAG	TTGCTGTGAG	AGGCTGAGAG	TTGCTGAGCT	TGAGGAGGAG	TATGCTGCTG	673095
TTGCTGTGAG	TTGCTGTGAG	AGGCTGAGAG	TTGCTGAGCT	TGAGGAGGAG	TATGCTGCTG	673100
TTGCTGTGAG	TTGCTGTGAG	AGGCTGAGAG	TTGCTGAGCT	TGAGGAGGAG	TATGCTGCTG	673105
TTGCTGTGAG	TTGCTGTGAG	AGGCTGAGAG	TTGCTGAGCT	TGAGGAGGAG	TATGCTGCTG	673110
TTGCTGTGAG	TTGCTGTGAG	AGGCTGAGAG	TTGCTGAGCT	TGAGGAGGAG	TATGCTGCTG	673115
TTGCTGTGAG	TTGCTGTGAG	AGGCTGAGAG	TTGCTGAGCT	TGAGGAGGAG	TATGCTGCTG	673120
TTGCTGTGAG	TTGCTGTGAG	AGGCTGAGAG	TTGCTGAGCT	TGAGGAGGAG	TATGCTGCTG	673125
TTGCTGTGAG	TTGCTGTGAG	AGGCTGAGAG	TTGCTGAGCT	TGAGGAGGAG	TATGCTGCTG	673130
TTGCTGTGAG	TTGCTGTGAG	AGGCTGAGAG	TTGCTGAGCT	TGAGGAGGAG	TATGCTGCTG	673135
TTGCTGTGAG	TTGCTGTGAG	AGGCTGAGAG	TTGCTGAGCT	TGAGGAGGAG	TATGCTGCTG	673140
TTGCTGTGAG	TTGCTGTGAG	AGGCTGAGAG	TTGCTGAGCT	TGAGGAGGAG	TATGCTGCTG	673145
TTGCTGTGAG	TTGCTGTGAG	AGGCTGAGAG	TTGCTGAGCT	TGAGGAGGAG	TATGCTGCTG	673150
TTGCTGTGAG	TTGCTGTGAG	AGGCTGAGAG	TTGCTGAGCT	TGAGGAGGAG	TATGCTGCTG	673155
TTGCTGTGAG	TTGCTGTGAG	AGGCTGAGAG	TTGCTGAGCT	TGAGGAGGAG	TATGCTGCTG	673160
TTGCTGTGAG	TTGCTGTGAG	AGGCTGAGAG	TTGCTGAGCT	TGAGGAGGAG	TATGCTGCTG	673165
TTGCTGTGAG	TTGCTGTGAG	AGGCTGAGAG	TTGCTGAGCT	TGAGGAGGAG	TATGCTGCTG	673170
TTGCTGTGAG	TTGCTGTGAG	AGGCTGAGAG	TTGCTGAGCT	TGAGGAGGAG	TATGCTGCTG	673175
TTGCTGTGAG	TTGCTGTGAG	AGGCTGAGAG	TTGCTGAGCT	TGAGGAGGAG	TATGCTGCTG	673180
TTGCTGTGAG	TTGCTGTGAG	AGGCTGAGAG	TTGCTGAGCT	TGAGGAGGAG	TATGCTGCTG	673185
TTGCTGTGAG	TTGCTGTGAG	AGGCTGAGAG	TTGCTGAGCT	TGAGGAGGAG	TATGCTGCTG	673190
TTGCTGTGAG	TTGCTGTGAG	AGGCTGAGAG	TTGCTGAGCT	TGAGGAGGAG	TATGCTGCTG	673195
TTGCTGTGAG	TTGCTGTGAG	AGGCTGAGAG	TTGCTGAGCT	TGAGGAGGAG	TATGCTGCTG	673200

[illegible]

FIGURE 5R

[illegible]

FIGURE 5S

GGGCTACGAA	GTGAGAGTGT	GTGTCGAAAA	ATGAAAAATRA	TATGATATAT	GATATGCGGA	75660
GTGCTATAGG	AGCTAGAGTG	TAAAAATGTAG	ATGAGACAGG	TGGGATTTTG	TGTTTCTCTT	75720
GTTCAGCGGT	AGGCTGGGAG	GTTCATTTTG	ATGATGAGGT	GTGCTGGGA	AGTGGCTTAC	75780
GTTCAGAAATG	GTGCTGCTCA	AGGCTGCTCT	GTGCTGAAAA	TAGGATTTTG	AGGAAAAAGT	75840
GTGCTGCTTG	GTGCTGATTT	TGGTGTGGGG	GGATGTATAT	TAGGCTTTTA	GTGATTTGAG	75900
GTGCTGCTGA	AACTTGAAGG	GAGGAGCTGA	GTGCTGAGCTG	GTGAGATATA	GAATTTGGAA	75960
GGATTTAGAT	GTGATGTGG	TTGAGAGGCT	GGGCTTCTCT	GGGACTTTCT	TGTTGTCTGT	76020
GAGTTTACAA	GTGCTGCTTA	TTGTTAGTCT	GTGCTGATGG	GTGCTGATCT	AGTGTGAGTC	76080
AAAACTTTTA	TTAGGCGCATG	ATGCTACTTTA	TTGCTATCTT	GTGCTGCTCT	TATTACAGCT	76140
ATATTTACAA	GTGCTGCTTG	ATTGAGCTCT	TGATTTTAA	GAGGCTGAG	AACAGCTCTA	76200
TGCTAGAGT	TGCTGAGTGG	TTGTTTAAATG	GGAGGCTGCTG	TTGTTGCTCT	TGAAATCTTA	76260
GTGCTGCTTG	TAGCTGCTAG	ATGCTATCTT	TTTAAAAAT	TATTTTCTCT	TTAGGCTGAG	76320
GATCTGCTTG	TGCTGCTGAG	GTGCTGCTTG	AGTGGCTATTA	TGATGCTCTA	TTGCTGCTCT	76380
AAAGCTGCTTG	GTGCTGCTGA	TGCTGCTGCT	TGCTGCTGCT	TAAGTGGCTG	GATTACAGCT	76440
GTGCTGCTTA	GGGCTGCTTG	TATTATTTCT	GATTCGAGAT	TACAGATGAG	GAATTAAGG	76500
GTTCAGGAGG	GTGCTGATTT	TGCTGAGATG	GTGCTGCTGCT	TAGGCTGCTG	GTGCTGCTTG	76560
AAAGCTGCTG	GTGCTGCTG	GTGCTGCTG	TGCTGCTGCT	TGCTGCTGCT	GTGCTGCTG	76620
GTGCTGCTTG	ATGCTGCTG	ATGCTGCTG	TGCTGCTGCT	GTGCTGCTG	GTGCTGCTG	76680
GTGCTGCTTG	GTGCTGCTG	GAGGATTTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	76740
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	76800
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	76860
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	76920
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	76980
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	77040
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	77100
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	77160
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	77220
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	77280
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	77340
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	77400
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	77460
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	77520
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	77580
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	77640
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	77700
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	77760
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	77820
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	77880
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	77940
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	78000
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	78060
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	78120
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	78180
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	78240
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	78300
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	78360
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	78420
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	78480
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	78540
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	78600
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	78660
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	78720
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	78780
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	78840
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	78900
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	78960
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	79020
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	79080
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	79140
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	79200
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	79260
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	79320
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	79380
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	79440
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	79500
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	79560
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	79620
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	79680
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	79740
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	79800
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	79860
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	79920
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	79980
GTGCTGCTTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	GTGCTGCTG	80040

FIGURE 5T

[illegible]

FIGURE 5U

TAGAGAGCTTT	TTGTGCGGAAA	AAGCGTGTGTT	TAGCTAGAGA	TTTGCGGTGT	ATGTAAGCTT	34060
AAAAATATTA	AGATTAAGAAA	ATTAAGAGAAA	TAGAGCTTGG	ATGGAATTTT	TTTGCGGAATG	34120
AATCTAGAAA	AATAAGAAAT	GATTGAAAT	AATGTAATAG	GAAGGCTGTA	TAAAAATGAT	34180
TTAGGGTTTA	TTAGGTTGAT	TTAGATATAT	ATGCTGTAAG	AAAGATATGA	AGGGGTATAA	34240
TTAGAGATAT	AAATAAATGG	ATTAAGATGA	AACATATTA	ATTAAGATAT	ATCTAATAAA	34300
AAAAAACTTA	GGAGTTTGG	TAGAGAGATC	GGAGAGATAT	TTTCTGAT	AGTAAGGAAT	34360
TAGCTATAT	TTTAAATAT	GGAAAAAGTT	AGAAAAAGG	AGCTTAATGA	CTGAATATTA	34420
AGTTTCTGTT	AACTTTTCTT	TGATATCTTT	TGAAGGCTAT	TTTAAAGAT	ATATATTTAA	34480
TAAGATATGAT	TTGATGTTTA	GAATTTGAGT	TGATTTTCTT	TTGCTTAAT	TGTAATGAAA	34540
TTAAATTAAT	AGCTTTTAA	AAAAATGGTT	TGATCTGAT	TAAATTAAT	TTCTTTTCTT	34600
TAGCTATATTA	TGATATATTA	ATCTGATAT	ATCTGATAT	GAATGGATAC	ATGTTGATAC	34660
ATTAATGGGG	TTTCTGTTTA	AGGGCTGAGA	GGGTATGTTT	TTCTAACCTT	AACCTGTGAC	34720
AGGCTGAGG	AGGGTAATGG	AATCTGTTT	AAATATTTT	AGCAAGAGAG	GGAGACAGAC	34780
AAAGATGGGG	TTTACACATC	GGTTTGGGAG	TGAGAGAGGT	GAAGAAAAAGG	GGATATGACT	34840
GGTCTGGTGA	TACAGACAGGG	AGGGTGGGCTG	GGGTGCTGGG	TGGGGCTGCT	CTTCTCTAGAG	34900
GAATCTGGGG	AAAGGGATAT	GGGGTGATTT	GGTTTGAGG	CATTATAGCC	AGAGTAGGGGA	34960
AGGTTTCTCT	ATGGGGTCTT	TTGGGGCTTT	CTTGACAGCG	CTTGACAGCG	AGAGTAGGGGA	35020
TGCGGAGAGA	TTGAAGGAGG	AGTGGGTGAG	AGGTTTCTTA	GGTCTAGGGA	AGCTTCTGAT	35080
TTCTGATTA	AGGCTGATC	ATTAATTTCT	TGCAATGTTT	AGCGAGAGCT	AGCTTGGGAT	35140
TTAGAGGAG	TAAATATCTT	TTTCTTCTT	AGTGGCTT	AGTGGCTT	GAGATATGAG	35200
AGGTCTTGAA	TTGTAGCTTT	CGATGGGGG	AGAAATAGGC	TGGGAGAGAG	AGAGTAGGAG	35260
AGGAACTGAA	GGGTCTTTCT	GGAGCTTGGA	AGAGAAAGGA	GTAAAAATG	AGGGATGGAT	35320
AAAAATATTA	GGCTAAATTA	AGTTTGGGAA	TTTCTGAGG	ATTAAGGCTT	AGCTGAGCTT	35380
GAATCTGAG	TTGCTTAAGAA	ATCTGATAT	TTGATATGCT	TGAGATATG	GGATATGAGG	35440
GAATCTGAG	TAAGCTGAGG	TGCTGGGGCT	TGCTGGGGCT	GTAGAGATG	TGATTTCTTT	35500
TTTCTGAG	TAAGCTGAG	AAATCTGAG	GTGAGAGGA	TTTCTGAGG	TGATTTCTTT	35560
TAAGCTGAG	TAATATGCTT	TTAGAGGCTT	TTGAGGCTT	AGCTTAAGGA	TTTCTGAGCT	35620
TAGATATG	TAGATATGCT	TAATATGCTT	ATTAATGCTT	AGAGAGAGT	TGATCTCTTT	35680
TGATCTGAG	AGCTTTGCTT	TAGATATGCT	AGAGAGAGT	AGAGAGAGT	TTTCTGAGCT	35740
AGCTTTGCT	TTTCTGAGG	TTTCTGAGG	GGAGAGATTA	GTAGATATG	TGATCTCTTT	35800
TTTCTGAG	ATTAATGCTT	ATTAATGCTT	TTTCTGAGG	GGAGAGATTA	TAGCTTTCTT	35860
AGCTTTGCT	ATTAATGCTT	TTTCTGAGG	AGAGATATG	GGAGAGATTA	TTTCTGAGCT	35920
AGCTTTGCT	ATTAATGCTT	TTTCTGAGG	AGAGATATG	GGAGAGATTA	TTTCTGAGCT	35980
AGCTTTGCT	ATTAATGCTT	TTTCTGAGG	AGAGATATG	GGAGAGATTA	TTTCTGAGCT	36040
AGCTTTGCT	ATTAATGCTT	TTTCTGAGG	AGAGATATG	GGAGAGATTA	TTTCTGAGCT	36100
AGCTTTGCT	ATTAATGCTT	TTTCTGAGG	AGAGATATG	GGAGAGATTA	TTTCTGAGCT	36160
AGCTTTGCT	ATTAATGCTT	TTTCTGAGG	AGAGATATG	GGAGAGATTA	TTTCTGAGCT	36220
AGCTTTGCT	ATTAATGCTT	TTTCTGAGG	AGAGATATG	GGAGAGATTA	TTTCTGAGCT	36280
AGCTTTGCT	ATTAATGCTT	TTTCTGAGG	AGAGATATG	GGAGAGATTA	TTTCTGAGCT	36340
AGCTTTGCT	ATTAATGCTT	TTTCTGAGG	AGAGATATG	GGAGAGATTA	TTTCTGAGCT	36400
AGCTTTGCT	ATTAATGCTT	TTTCTGAGG	AGAGATATG	GGAGAGATTA	TTTCTGAGCT	36460
AGCTTTGCT	ATTAATGCTT	TTTCTGAGG	AGAGATATG	GGAGAGATTA	TTTCTGAGCT	36520
AGCTTTGCT	ATTAATGCTT	TTTCTGAGG	AGAGATATG	GGAGAGATTA	TTTCTGAGCT	36580
AGCTTTGCT	ATTAATGCTT	TTTCTGAGG	AGAGATATG	GGAGAGATTA	TTTCTGAGCT	36640
AGCTTTGCT	ATTAATGCTT	TTTCTGAGG	AGAGATATG	GGAGAGATTA	TTTCTGAGCT	36700
AGCTTTGCT	ATTAATGCTT	TTTCTGAGG	AGAGATATG	GGAGAGATTA	TTTCTGAGCT	36760
AGCTTTGCT	ATTAATGCTT	TTTCTGAGG	AGAGATATG	GGAGAGATTA	TTTCTGAGCT	36820
AGCTTTGCT	ATTAATGCTT	TTTCTGAGG	AGAGATATG	GGAGAGATTA	TTTCTGAGCT	36880
AGCTTTGCT	ATTAATGCTT	TTTCTGAGG	AGAGATATG	GGAGAGATTA	TTTCTGAGCT	36940
AGCTTTGCT	ATTAATGCTT	TTTCTGAGG	AGAGATATG	GGAGAGATTA	TTTCTGAGCT	37000
AGCTTTGCT	ATTAATGCTT	TTTCTGAGG	AGAGATATG	GGAGAGATTA	TTTCTGAGCT	37060
AGCTTTGCT	ATTAATGCTT	TTTCTGAGG	AGAGATATG	GGAGAGATTA	TTTCTGAGCT	37120
AGCTTTGCT	ATTAATGCTT	TTTCTGAGG	AGAGATATG	GGAGAGATTA	TTTCTGAGCT	37180

FIGURE 6

GTATAAAGTT AGTAAATGTS AGGCCTCTCT CGATGCCTGG GTCTTGGGCT TTGGTTCTCA	60
GTCTCTCCTTA AATCATCTCT CTGGAGGAGA AGACCTTAG ATCTGGCTCT TCTCAGGGGC	120
ATTTTAAAGA CAAATGAAAA TAAA ATG GAA ACC ACT TCA CTA CAG CGG AAA	171
Met Glu Thr Thr Ser Leu Gln Arg Lys	
1 5	
TTT CCA GAA TGG ATG TCT ATG CAG AGT CAA AGA TGT GCT ACA GAA GAA	219
Phe Pro Glu Trp Met Ser Met Gln Ser Gln Arg Cys Ala Thr Glu Glu	
10 15 20 25	
AAG GCG TGG GTT CAG AAG AGT GTT CTT GAA GAT AAG CTC GCA TTC TTA	267
Lys Ala Cys Val Gln Lys Ser Val Leu Glu Asp Asn Leu Pro Phe Leu	
30 35 40	
GAA TTC CTT GGA TCC ATT GTT TAC AGT TAT GAA GCT AGT GAT TGC TCC	315
Glu Phe Pro Gly Ser Ile Val Tyr Ser Tyr Glu Ala Ser Asp Cys Ser	
45 50 55	
TTC CTC TCT GAA GAC ATT AGC ATG CTT CTC TCT GAT GGC GAT CTC GTC	363
Phe Leu Ser Glu Asp Ile Ser Met Arg Leu Ser Asp Gly Asp Val Val	
60 65 70	
GGA TTT GAC ATG GAA TGG TCC GCG ATA TAC AAG CCA GGG AAA AGA AGC	411
Gly Phe Asp Met Glu Trp Pro Pro Ile Tyr Lys Pro Gly Lys Arg Ser	
75 80 85	
AGA CTC GCA CTC ATT CAG TTC TGT CTC TCT GAG AGC AAA TGT TAC TTG	459
Arg Val Ala Val Ile Gln Leu Cys Val Ser Glu Ser Lys Cys Tyr Leu	
90 95 100 105	
TTT CAG ATT TCT TCC ATG TCA GTT TTC CCG CAG GGA TTA AAA ATG TTA	507
Phe His Ile Ser Ser Met Ser Val Phe Pro Gln Gly Leu Lys Met Leu	
110 115 120	
CTA GAA AAG AAA TCA ATT AAG AAG GCA GGG GTT GGG ATT GAA GGG GAC	555
Leu Glu Asn Lys Ser Ile Lys Lys Ala Gly Val Gly Ile Glu Gly Asp	
125 130 135	
CAG TCG AAA CTT CTC CTT GAT TTT GAC CTC AAG TTG GAG AGT TTT CTC	603
Gln Trp Lys Leu Leu Arg Asp Phe Asp Val Lys Leu Glu Ser Phe Val	
140 145 150	
GAG CTC ACG GAT GTT GCG AAT GAA AAG TTG AAG TCC GCA GAG ACC TGG	651
Glu Leu Thr Asp Val Ala Asn Glu Lys Leu Lys Cys Ala Glu Thr Trp	
155 160 165	
AGC CTC AAT GGT CTC GTT AAA CAC GTC TTA GCG AAA CAA TTT TTG AAA	699
Arg Leu Asn Gly Thr Val Lys His GTC TTA GCG AAA CAA TTT TTG AAA	
170 175 180	

FIGURE 6 (CONT.)

GAC CAG AAA CTG TAT GCA GCG ACT GAT GCT TAT GGT GGT GTT ATC ATC Asp Gln Lys Leu Tyr Ala Ala Thr Asp Ala Tyr Ala Gly Leu Ile Ile 205 210 215	795
TAT CAA AAA TTA GGA AAT TTG GGT GAT ACT GCG CAA GTG TTT GCT CTA Tyr Gln Lys Leu Gly Asn Leu Gly Asp Thr Ala Gln Val Phe Ala Leu 220 225 230	843
AAT AAA GCA GAG GAA AAC CTA CCT CTG GAG ATG AAG AAA CAG TTG AAT Asn Lys Ala Glu Glu Asn Leu Pro Leu Glu Met Lys Lys Gln Leu Asn 235 240 245	891
TCA ATC TCG GAA GAA ATG AGG GAC CTA GCG AAT GGT TTT CCT GTC ACT Ser Ile Ser Glu Glu Met Arg Asp Leu Ala Asn Arg Phe Pro Val Thr 250 255 260 265	939
TGC AGA AAT TTG GAA ACT CTC GAG AGG GTT CCT GTA ATA TTG AAG AGT Cys Arg Asn Leu Glu Thr Leu Gln Arg Val Pro Val Ile Leu Lys Ser 270 275 280	987
ATT TCA GAA AAT CTC TGT TCA TTG AGA AAA GTG ATC TGT GGT CCT AGA Ile Ser Glu Asn Leu Cys Ser Leu Arg Lys Val Ile Cys Gly Pro Thr 285 290 295	1035
AAC ACT GAG ACT AGA CTC ACG CCG GCG AGT AGT TTT AAT TTA CTC TCA Asn Thr Glu Thr Arg Leu Lys Pro Gly Ser Ser Phe Asn Leu Leu Ser 300 305 310	1083
TCA GAG GAT TCA GGT GGT GGT GGA GAA AAA GAG AAA CAG ATT GGA AAA Ser Glu Asp Ser Ala Ala Ala Gly Glu Lys Glu Lys Gln Ile Gly Lys 315 320 325	1131
CAT AGT ACT TTT GGT AAA ATT AAA GAA GAA CCA TGG GAC CCA GAA GTT His Ser Thr Phe Ala Lys Ile Lys Glu Glu Pro Trp Asp Pro Glu Leu 330 335 340 345	1179
GAC ACT TTA GTT AAG CAA GAG GAG GTT GAT GTA TTT AGA AAT GAA GTG Asp Ser Leu Val Lys Gln Glu Glu Val Asp Val Phe Arg Asn Gln Val 350 355 360	1227
AAG CAA AAA AAA GGT GAA TGT GAA AAT GAA ATA GAA GAC AAT CTG TTG Lys Gln Glu Lys Gly Glu Ser Glu Asn Glu Ile Glu Asp Asn Leu Leu 365 370 375	1275
AGA GAA GAT ATG GAA AGA ACT TGT GTG ATT GGT AGT ATT TCA GAA AAT Arg Glu Asp Met Glu Arg Thr Cys Val Ile Pro Ser Ile Ser Glu Asn 380 385 390	1323
GAA GTT CAA GAT TTG GAA TAA CAA GGT AAA GAA GAA AAA TAT GAT GAT Glu Val Cys Asp TTG GAA TAA CAA GGT AAA GAA GAA AAA TAT GAT GAT 395 400 405	1371

FIGURE 6 (CONT.)

GAC	TCC	TCC	TAT	ATA	ATT	GAA	AGT	GAT	GAA	GAT	TTG	GAA	ATG	GAG	ATG	1467
Asp	Ser	Ser	Tyr	Ile	Ile	Glu	Ser	Asp	Glu	Asp	Leu	Glu	Met	Glu	Met	
				430					435					440		
CTG	AAG	TCT	TTA	GAA	AAC	CTA	AAT	AGT	GAC	GTG	GTG	GAA	CCC	ACT	CAC	1515
Leu	Lys	Ser	Leu	Glu	Asn	Leu	Asn	Ser	Asp	Val	Val	Glu	Pro	Thr	His	
			445					450					455			
TCT	ACA	TGG	TTG	GAA	ATG	GGA	ACC	AAT	GGG	CGT	CTT	CCT	ECT	GAG	GAG	1563
Ser	Thr	Trp	Leu	Glu	Met	Gly	Thr	Asn	Gly	Arg	Leu	Pro	Pro	Glu	Glu	
		460					465					470				
GAA	GAT	GGA	CAC	GGA	AAT	GAA	GCC	ATC	AAA	GAG	GAG	CAG	GAA	GAA	GAG	1611
Glu	Asp	Gly	His	Glv	Asn	Glu	Ala	Ile	Lys	Glu	Glu	Gln	Glu	Glu	Glu	
	475					480						485				
GAC	CAT	TTA	TTG	CCC	GAA	CCC	AAC	GCA	AAG	GAA	ATT	AAT	TGC	CTC	AAG	1659
Asp	His	Leu	Leu	Pro	Glu	Pro	Asn	Ala	Lys	Gln	Ile	Asn	Cys	Leu	Lys	
490				495					500				505			
ACC	TAT	TTG	GGA	CAC	AGC	AGT	TTT	AAA	CCC	GTT	CAG	TGC	AAA	CTC	ATC	1707
Thr	Tyr	Phe	Gly	His	Ser	Ser	Phe	Lys	Pro	Val	Gln	Trp	Lys	Val	Ile	
			510					515					520			
TAT	TCT	GTA	TTA	GAA	GAG	AGA	AGA	GAT	AAT	GTT	GTT	GTC	ATG	GCA	ACT	1755
His	Ser	Val	Leu	Glu	Glu	Arg	Arg	Asp	Asn	Val	Val	Val	Met	Ala	Thr	
			525					530					535			
GGA	TAT	GGG	AAG	ACT	CTG	TGG	TTG	CAG	TAT	CCG	CGT	GTT	TAT	ACA	GCG	1803
Gly	Thr	Gly	Lys	Ser	Leu	Cys	Phe	Gln	Tyr	Pro	Phe	Val	Tyr	Thr	Gly	
		540				545						550				
AAG	ATT	GGC	ATT	GTC	ATT	TCA	CGT	CTC	ATT	TCC	TTA	ATG	GAA	GAC	CAA	1851
Lys	Ile	Gly	Ile	Val	Ile	Ser	Pro	Leu	Ile	Ser			Glu	Asp	Gln	
	555					560				565						
GTC	GTC	GGG	TTT	GAG	CTG	TCC	AAT	GTT	GCA	GGC	TGT	TTA	GTT	GGA	TCT	1899
Val	Leu	Gln	Leu	Glu	Leu	Ser	Asn	Val	Pro	Ala	Cys	Leu	Leu	Gly	Ser	
570				575					580				585			
GCA	CAG	TCA	AAA	AAT	ATT	CTA	GGA	GAT	CTT	AAA	TTA	GGC	AAA	TAT	AGG	1947
Ala	Gln	Ser	Lys	Asn	Ile	Leu	Gly	Asp	Val	Lys	Leu	Gly	Lys	Tyr	Arg	
			590					595					600			
GTC	ATC	TAC	ATA	ACT	GCA	GAG	TTC	TGT	TCT	GGT	AAC	TTG	GAT	CTA	CTC	1995
Val	Ile	Tyr	Ile	Thr	Pro	Glu	Phe	Cys	Ser	Gly	Asn	Leu	Asp	Leu	Leu	
			605					610					615			
GAG	GAA	TTT	GAG	TCT	ACT	AAT	CTC	ATT	ACT	CTC	ATT	CTT	GTC	GAT	GAA	

FIGURE 6 (CONT.)

ATG CTG GGC TGT GTT AAA ACA GCG CTC GCA TTG GTT CCA GTC ATT GCA	2139
Met Leu Gly Ser Leu Lys Thr Ala Leu Pro Leu Val Pro Val Ile Ala	
650 655 660 665	
CTC TCC GGT ACT GCA AGC TCT TCC ATC CGG GAA GAC ATT ATA AGC TGC	2187
Leu Ser Ala Thr Ala Ser Ser Ser Ile Arg Glu Asp Ile Ile Ser Cys	
670 675 680	
TTA AAC CTG AAA GAC GGT CAG ATC ACC TGC ACT GGA TTT GAT CGG CCA	2235
Leu Asn Leu Lys Asp Pro Gln Ile Thr Cys Thr Gly Phe Asp Arg Pro	
685 690 695	
AAT CTG TAC TTA GAA GTT GGA CGG AAA ACA GGG AAC ATC CTT CAG GAT	2283
Asn Leu Tyr Leu Glu Val Gly Arg Lys Thr Gly Asn Ile Leu Gln Asp	
700 705 710	
ATA AAG CCG TTT GTT CTC CGA AAG GCA AGT TCT GGC TGG GAA TTT GAA	2331
Leu Lys Pro Phe Leu Val Arg Lys Ala Ser Ser Ala Trp Ile Phe Gly	
715 720 725	
GST CCA AGC ATC ATC TAT TGT GGT TCG AGA AAA ATG ACA GAA CAA GTT	2379
Gly Pro Thr Ile Ile Tyr Cys Pro Ser Arg Lys Met Thr Gln Gln Val	
730 735 740 745	
ACT GGT GAA GTT GGG AAA CTG AAC TTA GCG TGG AGA ACA TAC CAC GGT	2427
Thr Ala Gly Leu Gly Lys Leu Asn Leu Ala Cys Arg Thr Tyr His Ala	
750 755 760	
GGC ATG AAA ATT ACC GAA AGG AAG GAC GTT CAT CAT AGG TTC CTG AGA	2475
Gly Met Lys Ile Ser Glu Arg Lys Asp Val His His Arg Phe Leu Arg	
765 770 775	
GAT GAA ATT CAG TGT GTT GTA GGT ACT GTA GGT TTT GGA ATG GGC ATT	2523
Asp Glu Ile Gln Cys Val Val Ala Thr Val Ala Phe Gly Met Gly Ile	
780 785 790	
AAT AAA GGT GAC ATT TCG AAA GTT ATT CAT TAT GGT GCG GGT AAG GAA	2571
Asn Lys Ala Asp Ile Arg Lys Val Ile His Tyr Gly Ala Pro Lys Gly	
795 800 805	
ATG GAA TCC TAT TAC CAG GAA ATT GST AGA GGT GGC CGG GAT GGA GTT	2619
Met Glu Ser Tyr Tyr Gln Glu Ile Gly Arg Ala Gly Arg Asp Gly Leu	
810 815 820 825	
CAG AGT TCC TGT CAC TTG CTC TGG GGT CCA GCA GAC TTT AAC AAG TCC	2667
Gln Ser Ser Cys His Leu Leu Trp Ala Pro Ala Asp Phe Asn Thr Ser	
830 835 840	
AGG AAT CTC GTT ATT GAG ATT CAG GAT GAA AAG TTC TGG TTA TAT AAA	2715

FIGURE 6 (CONT.)

AGG CGA CGA ATC ATC TTG TCC CAT TTT GAG GAC AAA TGT CTG CAG AAG	2811
Arg Arg Arg Ile Ile Leu Ser His Phe Glu Asp Lys Cys Leu Gln Lys	
375 880 885	
GCC TCC TTG GAC ATT ATG GGA ACT GAA AAA TGC TGT GAT AAT TCC AGG	2859
Ala Ser Leu Asp Ile Met Gly Thr Glu Lys Cys Cys Asp Asn Cys Arg	
890 895 900 905	
CCC AGG CTG AAT CAT TGC ATT ACT GCT AAC AAC TCA GAG GAC GCA TCC	2907
Pro Arg Leu Asn His Cys Ile Thr Ala Asn Asn Ser Glu Asp Ala Ser	
910 915 920	
CAA GAC TTT GGG CCA CAA GCA TTC CAG CTA CTG TCT GCT GTG GAC ATC	2955
Gln Asp Phe Gly Pro Gln Ala Phe Gln Leu Leu Ser Ala Val Asp Ile	
925 930 935	
CTG CAG GAG AAA TTT GGA ATT GGG ATT CCG ATC TTA TTT CTC CCA GGA	3003
Leu Gln Gln Lys Phe Glu Ile Gly Ile Pro Ile Leu Phe Leu Arg Gly	
940 945 950	
TCT AAT TCT CAG CTT CTT CTT GAT AAA TAT CCG GGT CAC AGG CTC TTT	3051
Ser Asn Ser Gln Arg Leu Pro Asp Lys Tyr Arg Gly His Arg Leu Phe	
955 960 965	
GCT GCT GCA AAG GAG CAA GTT GAA AGT TCG TCG AAG ACC CTT TCT CAC	3099
Gly Ala Gly Cys Gln Gln Ala Gln Ser Trp Trp Lys Thr Leu Ser His	
970 975 980 985	
CAT CTC ATA GCT GAA GGA TTC TTG GTA GAA GTT CCG AAG GCA AAC AAA	3147
His Leu Ile Ala Gln Gly Phe Leu Val Glu Val Pro Lys Glu Asn Lys	
990 995 1000	
TAT ATA AAG ACA TGT TCC CTC ACA AAA AAG GGT AGA AAG TGG CTT GGA	3195
Tyr Ile Lys Thr Cys Ser Leu Thr Lys Lys Gly Arg Lys Trp Leu Gly	
1005 1010 1015	
GAA GCG AGT TCG CAG TCT CTT CCG AGC CTT CTC CTT CAA GTT AAT GAA	3243
Glu Ala Ser Ser Gln Ser Pro Pro Ser Leu Leu Leu Gln Ala Asn Glu	
1020 1025 1030	
GAG ATG TTT TCA AGG AAA GTT CTG CTA CCA AGT TGT AAT CTT GTA TGT	3291
Glu Met Phe Pro Arg Lys Val Leu Leu Pro Ser Ser Asn Pro Val Ser	
1035 1040 1045	
CCA GAA AGG AGG CAA CAT TCC TCT AAT CAA AAC CCA GCT GGA TTA ACT	3339
Pro Glu Thr Thr Gln His Ser Ser Asn Gln Asn Pro Ala Gly Leu Thr	
1050 1055 1060 1065	
ACC AAG CAG TTT AAT TTG GAG AGA AGC CAT TTT TAC AAA GTG CTT CAG	3387
Thr Lys Gln Phe Asn Cys GAG AGA AGC CAT TTT TAC AAA GTG CTT CAG	

FIGURE 6 (CONT.)

TCA CCA GGA ACA TCT TCC AGC CCC TTA GAA CCT GCC ATC TCA GCC CAA Ser Pro Gly Thr Ser Ser Ser Pro Leu Glu Pro Ala Ile Ser Ala Gln 1100 1105 1110	3483
GAG CTC GAC GGT CCG ACT GGG CTA TAT GCC AGG CTG GTG GAA GCA AGG Glu Leu Asp Ala Arg Thr Gly Leu Tyr Ala Arg Leu Val Glu Ala Arg 1115 1120 1125	3531
CAG AAA CAC GGT AAT AAG ATG GAT GTA CCT CCA GCT ATT TTA GCA ACA Gln Lys His Ala Asn Lys Met Asp Val Pro Pro Ala Ile Leu Ala Thr 1130 1135 1140 1145	3579
AAC AAG GTT CTG CTG GAC ATG GCT AAA ATG AGA CCC ACT ACT GTT GAA Asn Lys Val Leu Leu Asp Met Ala Lys Met Arg Pro Thr Thr Val Glu 1150 1155 1160	3627
AAC ATG AAA CAG ATC GAC GGT GTG TCT GAA GCC AAA GGT GCT CTG TTG Asn Met Lys Gln Ile Asp Gly Val Ser Glu Gly Lys Ala Ala Leu Leu 1165 1170 1175	3675
GCC GGT CTG TTC GAA CTC ATC AAA GAT TTC TGT CAA GTA ACT ACT GTT Ala Pro Leu Leu Glu Val Ile Lys His Phe Cys Gln Val Thr Ser Val 1180 1185 1190	3723
CAG AAA GAC CTC GTT TCC ACT GCC AAA GGT CAC AAG GAA CAG GAG AAA Gln Thr Asp Leu Leu Ser Ser Ala Lys Pro His Lys Gln Gln Glu Lys 1195 1200 1205	3771
AGT CAG CAG ATC GAA AAG AAA GAC TGC TCA CTC GCC CAG TCT GTG GCC Ser Gln Glu Met Glu Lys Lys Asp Cys Ser Leu Pro Gln Ser Val Ala 1210 1215 1220 1225	3819
CTC ACA TAC ACT CTA TTC CAG GAA AAG AAA ATG GCC TTA CAC AGC ATA Val Thr Tyr Thr Leu Phe Gln Glu Lys Lys Met Pro Leu His Ser Ile 1230 1235 1240	3867
GCT CAG AAC AGG CTC CTG GGT CTC ACA GCA GCC GGC ATG CAG TTA GCC Ala Glu Asn Arg Leu Leu Pro Leu Thr Ala Ala Gly Met His Leu Ala 1245 1250 1255	3915
CAG CCG CTG AAA GCC GCG TAC CCC CTG GAT ATG GAG CGA GGT GGC CTG Gln Ala Val Lys Ala Gly Tyr Pro Leu Asp Met Glu Arg Ala Gly Leu 1260 1265 1270	3963
ACC CCA GAG ACT TGG AAG ATT ATT ATG GAT GTC ATC CGA AAC GGT CCC Thr Pro Glu Thr Trp Lys Ile Ile Met Asp Val Ile Arg Asn Pro Pro 1275 1280 1285	4011
ATC AAC TCA GAT ATG TAT AAA GTT AAA CTC ATC AGA ATG TTA GTT GGT Ile Asn Thr GAT ATG TAT AAA GTT AAA CTC ATC AGA ATG TTA GTT GGT 1290 1295 1300	4059

FIGURE 6 (CONT.)

AGT GGT TCC GAC ACC AGA ACC GAG CCT CCT TGT GAT TCC AGC AGG AAG Ser Gly Ser Asp Ser Arg Thr Gln Pro Pro Cys Asp Ser Ser Arg Lys 1325 1330 1335	4155
AGG CGT TTC CCC AGC TCT GCA GAG AGT TGT GAG AGC TGT AAG GAG AGC Arg Arg Phe Pro Ser Ser Ala Glu Ser Cys Glu Ser Cys Lys Glu Ser 1340 1345 1350	4203
AAA GAG GCG GTC ACC GAG ACC AAG GCA TCA TCT TCA GAG TCA AAG AGA Lys Glu Ala Val Thr Glu Thr Lys Ala Ser Ser Ser Glu Ser Lys Arg 1355 1360 1365	4251
AAA TTA CCC GAG TGG TTT GCC AAA GGA AAT GTC CCC TCA GGT GAT ACC Lys Leu Pro Glu Trp Phe Ala Lys Gly Asn Val Pro Ser Ala Asp Thr 1370 1375 1380 1385	4299
GGC AGC TCA TCA TCA ATG GGC AAG ACC AAA AAG AAA GGT CTC TTT AGT Gly Ser Ser Ser Ser Met Ala Lys Thr Lys Lys Lys Gly Leu Phe Ser 1390 1395 1400	4347
TAAATGACAA AGGATGGAAC AATTTGTGTG TCGTACATCT TCATTGCTAT AAAGAATGAA 4407	
NAGAAATATT TTAAGCTGAA AATTATTGAA AGTCGAAAGT GAAGCTCACC TAAAGCTGGA 4467	
GCCATAGAGT GTTTAATTGTA CGCTTGGCAG TTGAGCTACA GTATCTGAAC GTTCTGAGAC 4527	
CCGGAGTGA GCATAGACTG TGAAGTCGGC TTGCTTTGCG ATTGCGTTCC GAACCGGTGT 4587	
CAGTGTGAG TTGAGTGTG TGTCTCTTG CAGCAGTGTG TGTGGAAT GGAGGCTGTG 4647	
TGCTTTGAG ATATAGAACA GATCAATAT TGCATAGGGA CAGATATGAA GATNCAGCG 4707	
GTCTTGTGT TCTTATGGAG ATGCGTGTAT GACATATCA GTGCACGAGC CAGGCCAGGG 4767	
AGACATGAG TTGATTTAA AAGG 4792	

Sequence

01459 01459

[illegible]

[illegible]

FIGURE 7 (CONT.)

[illegible]

[illegible]

FIGURE 7 (CONT.)

[illegible]

FIGURE 7 (CONT.)

[illegible]

FIGURE 7 (CONT.)

[illegible]

[illegible]

[illegible]

[illegible]

FIGURE 7 (CONT.)

[illegible]

FIGURE 7 (CONT.)

[illegible]

[illegible]

FIGURE 7 (CONT.)

[illegible]

FIGURE 7 (CONT.)

[illegible]

[illegible]

the 1990s, the number of people in the world who are undernourished has declined from 760 million to 600 million. The number of people who are malnourished has declined from 1.1 billion to 800 million. The number of people who are obese has increased from 100 million to 300 million. The number of people who are overweight has increased from 100 million to 300 million. The number of people who are obese and overweight has increased from 100 million to 300 million. The number of people who are obese and overweight has increased from 100 million to 300 million.

FIGURE 7 (CONT.)

[illegible]

FIGURE 7 (CONT.)

[illegible]

FIGURE 7 (CONT.)

[illegible]

[illegible]

FIGURE 7 (CONT.)

[illegible]

FIGURE 7 (CONT.)

[illegible]

[illegible]

[illegible]

[illegible]

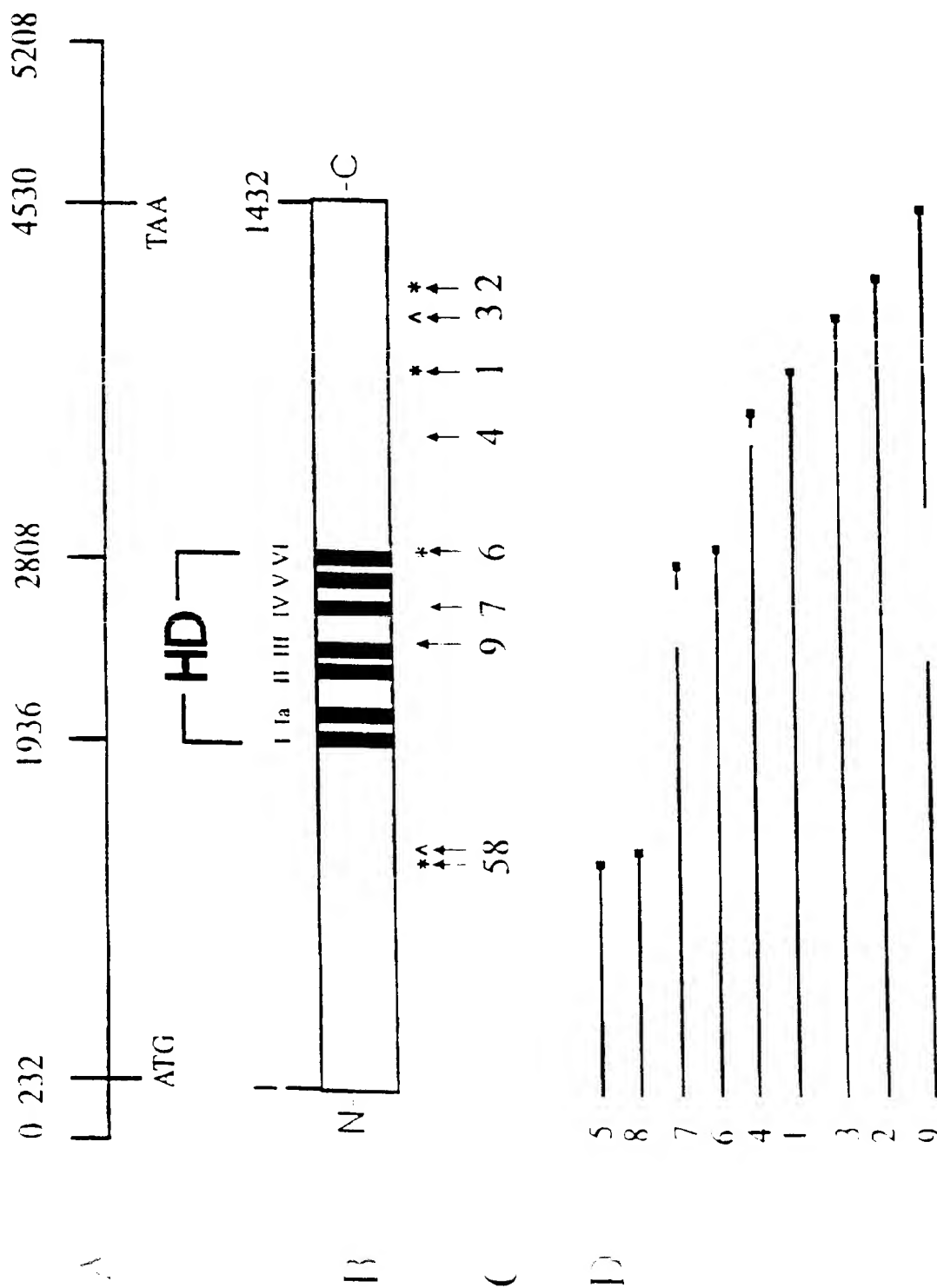


FIGURE 8

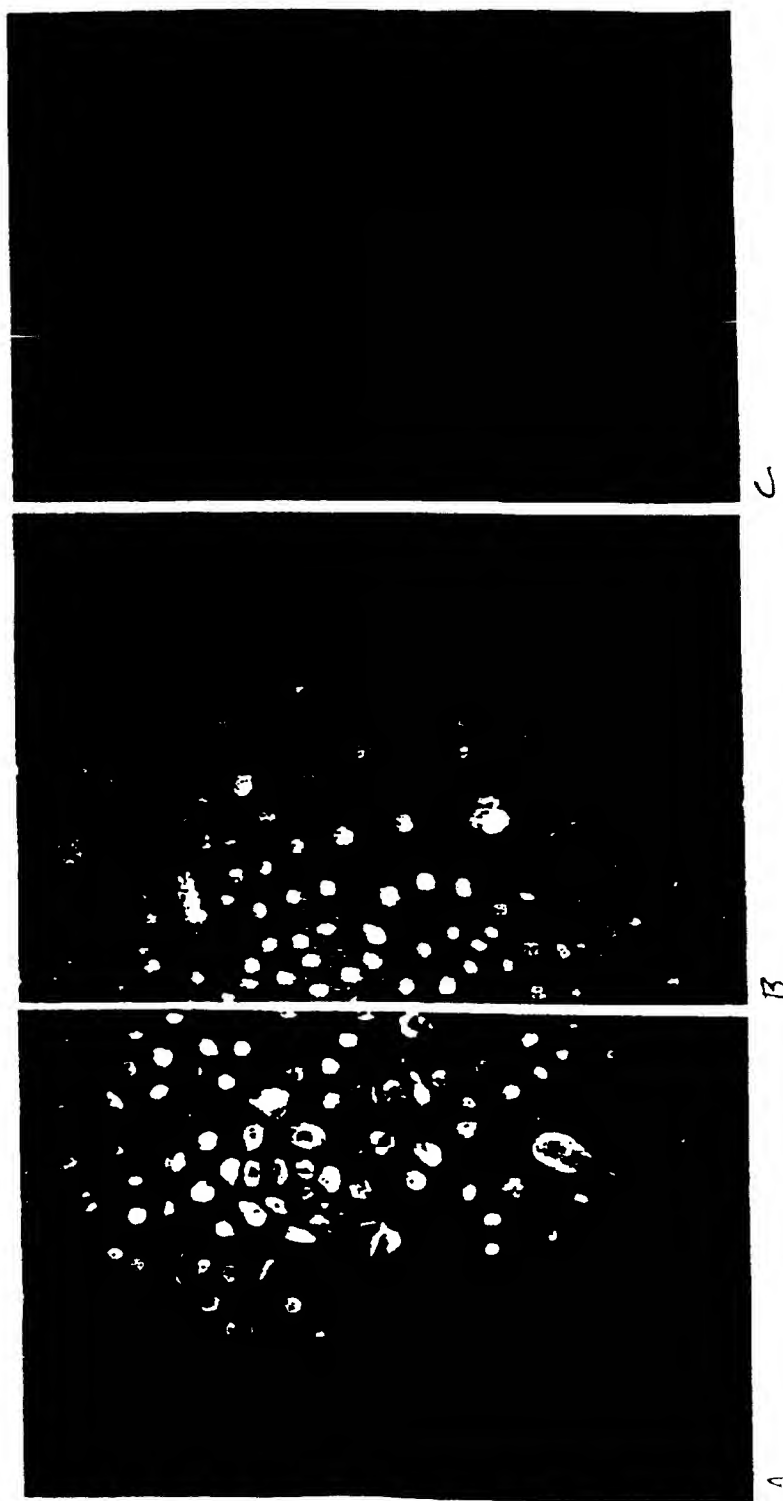


Figure 4. Expression of the WRN protein by indirect immunofluorescence in adherent human epithelial cells. Cells from prostate epithelium were placed on a glass slide and allowed to grow for several days before staining. The cells were fixed, permeabilized, and stained simultaneously for the WRN protein, DNA in the nucleus, or F-actin in the cytoplasm. Panel A shows cells stained for WRN (green) and DAPI (blue). Panel B shows cells stained for WRN (green) and F-actin (red). Panel C shows cells stained for WRN (green) and F-actin (red). Panel D shows cells stained for WRN (green) and F-actin (red). Panel E shows cells stained for WRN (green) and F-actin (red). Panel F shows cells stained for WRN (green) and F-actin (red). Panel G shows cells stained for WRN (green) and F-actin (red). Panel H shows cells stained for WRN (green) and F-actin (red). Panel I shows cells stained for WRN (green) and F-actin (red). Panel J shows cells stained for WRN (green) and F-actin (red). Panel K shows cells stained for WRN (green) and F-actin (red). Panel L shows cells stained for WRN (green) and F-actin (red). Panel M shows cells stained for WRN (green) and F-actin (red). Panel N shows cells stained for WRN (green) and F-actin (red). Panel O shows cells stained for WRN (green) and F-actin (red). Panel P shows cells stained for WRN (green) and F-actin (red). Panel Q shows cells stained for WRN (green) and F-actin (red). Panel R shows cells stained for WRN (green) and F-actin (red). Panel S shows cells stained for WRN (green) and F-actin (red). Panel T shows cells stained for WRN (green) and F-actin (red). Panel U shows cells stained for WRN (green) and F-actin (red). Panel V shows cells stained for WRN (green) and F-actin (red). Panel W shows cells stained for WRN (green) and F-actin (red). Panel X shows cells stained for WRN (green) and F-actin (red). Panel Y shows cells stained for WRN (green) and F-actin (red). Panel Z shows cells stained for WRN (green) and F-actin (red).

FIGURES 9A, B and C

0570 KAYUSIYAK STEEL 5/16" DIA. 1/2" L.

[illegible]

FIGURE 10 (CONTINUED)

Human	WVF GIVAFKILLSA VVLIAGEFGI GLEPIELIHS HSURLAKYVR	1050
Human	AKRS QAESWMTLS IILLJAEGLV EVFEEHYIK TQSLERGRK	1050
Human	TURD QTESWKAFS RQJITEGLV EVSRVIEFHE ICALTERGN	
Human	QSP FILLIQANEE MFRKVLVLS SRFVCHETQ IIRKQWPAH	1100
Human	RESL QULLIQANEE LAFKRFILPS RKFVREKTRD IIRVQVIVEL	
Human	QER TUSYKVEEV SSGNIPEKS AGRFSQESS SLEPAISAQ	1150
Human	QER LYSTKPCIKI SSGNISSEKS THUSPELAY SSSQVVISAO	
Human	YA NIVEARQKIA NENOVPPAL ATUKVELLIA KIRPTIVFIM	1200
Human	YU KIVEARQKIA NENOVPPAL ATUKVELLIA KIRPTIVFIV	
Human	EGR AALAPLEV IRIFCQTSV QTHLSSAKP IIRKQESQPM	1250
Human	EGR AALAPLEV IRIFCQTSV QTHLSSAKP IIRKQESQPM	
Human	US VAVTYTIVUE KEMPIJSIAE QHLLPTAAG IILLAQAVKAI	1300
Human	US HAITYSIVUE KEMPIJSIAE QHLLPTAAG IILLAQAVKAI	
Human	ACH TPEVKIIMD VIRNPINSQ HVEVKELIML VVEHIVTYLI	1350
Human	ACH TPEVKIIMD VIRNPINSQ HVEVKELIML VVEHIVTYLI	
Human	JRI GSRIVQPCD SSRKNEFSS AESSESSEKS KRAVTETRA	1400
Human	TRI FUSGLQFSCD VNRKNEFNS EETCSSSEKS KIRVGINLET	
Human	LP EMFAGNVPS ADTGSSSSMA KIRKKGLES*	1440
Human	QRP VNFAGS...DTSEKLMQ KIRKKGLES*	

Analysis and Last Update generated by A. Swartz

INTERNATIONAL SEARCH REPORT

International Application No.

PLT/US 96/20785

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/00 C12N15/61 C12N9/90 C12N15/85 C12N15/86
C12N5/10 C12N5/20 C07K16/40 C12Q1/68 A01K67/027
G01N33/50

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N C07K C12Q A01K G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EMBL SEQUENCE DATABASE, 30 January 1995, HEIDELBERG, BRD, XP002032088 F. BOUILLAUD: "Study of expressed sequence tags in adipose tissue 1995" accession no. T39125 see abstract	2,15,16
X	EMBL SEQUENCE DATABASE, 29 May 1995, HEIDELBERG, BRD, XP002032089 R. BERRY ET AL.: "Gene-based sequence tagged sites (STSs) as the basis for a human gene map" see accession no. R58879 see abstract	2,15,16

-/--

☒ Further documents are listed in the continuation of box C.

☐ Patent family members are listed in annex

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"I" document which may throw doubts on priority claims or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition, or other type of disclosure

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more prior art documents

Date of mailing of the international search report

2 June 1997

Date of mailing of the international search report

10.06.97

Name and mailing address of the ISA

European Patent Office, P.O. Box 1, 7000 Lausanne,
CH-1000, Switzerland
Tel: +41 21 314 1111
Fax: +41 21 314 1111

Authorized officer

ROBERT G. ...

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 96/20785

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>NATURE, vol. 355, 20 February 1992, MACMILLAN JOURNALS LTD., LONDON, UK, pages 735-738, XP002032090 M. GOTO ET AL.: "Genetic linkage of Werner's syndrome to five markers on chromosome 8" cited in the application see the whole document</p> <p>---</p>	1-21
A	<p>AMERICAN JOURNAL OF HUMAN GENETICS 55 (2). 1994. 356-364. ISSN: 0002-9297, August 1994, XP000674409 YU C-E ET AL: "Linkage disequilibrium and haplotype studies of chromosome 8p 11.1-2.1 markers and Werner syndrome." cited in the application see the whole document</p> <p>---</p>	1-21
A	<p>ANNUAL MEETING OF THE AMERICAN SOCIETY FOR BIOCHEMISTRY AND MOLECULAR BIOLOGY, SAN FRANCISCO, CALIFORNIA, USA, MAY 21-25, 1995. FASEB JOURNAL 9 (6). 1995. A1270. ISSN: 0892-6638, 24 April 1995, XP002032091 THWEATT R ET AL: "A novel cDNA overexpressed in Werner syndrome (WS) fibroblasts inhibits colony formation in normal human fibroblasts and in HeLa cells." abstract no.84 see abstract</p> <p>---</p>	1-21
A	<p>GENOMICS 28 (2). 1995. 147-153. ISSN: 0888-7543, 20 July 1995, XP000674353 KURIMASA A ET AL: "Construction of 110 cosmid markers and a 4.5-Mb YAC contig on human chromosome 8p12-q11." see the whole document</p> <p>---</p>	1-21
A	<p>GENOMICS 28 (3). 1995. 566-569. ISSN: 0888-7543, 10 August 1995, XP000674354 YE L ET AL: "Genetic association between chromosome 8 microsatellite (MS8-134) and Werner Syndrome (WRN): Chromosome microsatellite and human genetic markers</p>	1-21

INTERNATIONAL SEARCH REPORT

International Application No.
PCT/US 96/20785

C(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	45TH ANNUAL MEETING OF THE AMERICAN SOCIETY OF HUMAN GENETICS, MINNEAPOLIS, MINNESOTA, USA, OCTOBER 24-28, 1995. AMERICAN JOURNAL OF HUMAN GENETICS 57 (4 SUPPL.). 1995. A266. ISSN: 0002-9297, October 1995, XP000674410 NAKURA J ET AL: "Genetic association between chromosome 8 microsatellites and Werner syndrome (WRN)." abstract no. 1544 see abstract ---	1-21
A	NUCLEIC ACIDS RESEARCH, vol. 22, no. 22, 11 November 1994, IRL PRESS LIMITED, OXFORD, ENGLAND, pages 4566-4573, XP002032092 M. SEKI ET AL.: "Molecular cloning of cDNA encoding human DNA helicase Q1 which has homology to Escherichia coli Rec Q helicase and localization of the gene at chromosome 12p12" see the whole document ---	1-21
A	J. BIOL. CHEM., vol. 269, no. 47, 25 November 1994, AM. SOC. BIOCHEM. MOL. BIOL., INC., BALTIMORE, US, pages 29838-29845, XP002032093 K.L. PURANAM AND P.J. BLACKSHEAR: "Cloning and characterization of RECQL, a potential human homologue of the Escherichia coli DNA helicase RecQ" see the whole document ---	1-21
P,X	SCIENCE (WASHINGTON D C) 272 (5259). 1996. 258-262. ISSN: 0036-8075, XP002032094 YU C-E ET AL: "Positional cloning of the Werner's syndrome gene." see the whole document ---	1,2,10
P,X	TRENDS IN GENETICS 12 (8). 1996. 283-286. ISSN: 0168-9525, August 1996, XP002032095 LOMBARD D B ET AL: "Cloning the gene for Werner syndrome: A disease with many symptoms of premature aging." see the whole document ---	1,2,10, 15,16
P,X	HUMAN MOLECULAR GENETICS 5 (12). 1996 HUM. MOL. GEN. 5 (12): 1811-1815, 1996 heterozygous mutation of the WRN gene syndrome locus. see the whole document ---	1,2,10

INTERNATIONAL SEARCH REPORT

International Application No.
PCT/US 96/20785

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
T	AMERICAN JOURNAL OF HUMAN GENETICS 60 (2). 1997. 330-341. ISSN: 0002-9297, February 1997, XP000674411 YU C-E ET AL: "Mutations in the consensus helicase domains of the Werner syndrome gene." see the whole document ---	1-21
T	GENOMICS 41 (2). 1997. 298-300. ISSN: 0888-7543, XP000674356 IMAMURA O ET AL: "Cloning of a mouse homologue of the human Werner syndrome gene and assignment to 8A4 by fluorescence in situ hybridization." see the whole document -----	1-21

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US95/05330

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : A61K 48/00; C12N 5/00, 5/06, 5/08, 5/10, 5/12, 5/16, 5/22

US CL : 435/240.1, 240.2, 172.3, 320.1; 424/93.21; 935/62

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/240.1, 240.2, 172.3, 320.1; 424/93.21; 935/62

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CAS ONLINE, APS, bystander#, migration, msc, mc, gene#, dna#, cdna#, rna#, mrna#, gene, thera?

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 93/02556 [FREEMAN ET AL] 18 Februray 1993, see entire document.	1-33
Y	Science, Volume 256, issued 12 June 1992, K.W. Culver, "In vivo gene transfer with retroviral vector-producer cells for treatment of experimental brain tumors", pages 1550-1552, see entire document.	1-33
Y	Cancer Research, Volume 53, issued 01 January 1993, Z. Ram et al, "In situ retroviral-mediated gene transfer for the treatment of brain tumors in rats", pages 83-88, see entire document.	1-33

BEST AVAILABLE COPY

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
A document defining the general state of the art which is not considered to be of particular relevance	*N* document of the state of the art which is not considered to be of particular relevance
E earlier document published on or after the international filing date but earlier than the priority date	*G* document member of the same patent family
P document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

24 JULY 1995

Date of mailing of the international search report

03 AUG 1995

Name and mailing address of the ISA-US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

JAMES MARTINELL

Telephone No. (703) 308-0196

International application No
PCT/US95/05330

International application No
PCT/US95/05330

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Proceedings of the National Academy of Sciences, USA, Volume 85, issued October 1988, E. Borrelli et al, "Targeting of an inducible toxic phenotype in animal cells", pages 7572-7576, see entire document.	1-33
Y	Journal of Experimental Medicine, Volume 172, issued October 1990, B. Gansbacher et al, "Interleukin 2 gene transfer into tumor cells abrogates tumorigenicity and induces protective immunity", pages 1217-1224, see entire document.	1-33
Y	Science, Volume 254, issued 01 November 1991, P.T. Golumbek et al, "Treatment of established renal cancer by tumor cells engineered to secrete interleukin-4", pages 713-716, see entire document.	1-33
Y	Journal of Experimental Medicine, Volume 177, issued May 1993, L.E. Minasi et al, "The selective ablation of interleukin 2-producing cells isolated from transgenic mice", pages 1451-1459, see entire document.	1-33

BEST AVAILABLE COPY

BEST AVAILABLE COPY